Lancaster University
Management School

Economics Working Paper Series

2021/009

Is Business Formation Driven by Sentiment or Fundamentals?

Kim Kaivanto and Peng Zhang

The Department of Economics
Lancaster University Management School
Lancaster LA1 4YX
UK

LUMS home page: http://www.lancaster.ac.uk/lums/

Is Business Formation Driven by Sentiment or Fundamentals?*

Kim Kaivanto[§] and Peng Zhang[⋆]

[§]Lancaster University, Lancaster LA1 4YX, UK

[⋆]College of Humanities and Sciences of Guizhou Minzu University, Guiyang, PRC 550025

this version: June 13, 2021

Abstract

The creation of a new business is an act of entrepreneurship. It is also a financial undertaking. Hence it is admissible to apply the apparatus of behavioral finance to study the determinants of business formation. Our results show that aggregate business formation, nationally and regionally, is jointly predicted by economic fundamentals and sentiment. There is evidence of both 'pull' and 'push' motives for entrepreneurship. Yet this simple structure does not survive decomposition by payroll propensity. High-payroll-propensity entrepreneurs respond primarily to pull-motive fundamentals, with sentiment accounting for a small fraction of explained variance. Low-payroll-propensity entrepreneurs, on the other hand, respond to both sentiment and fundamentals, representing both pull- and push-motives, with sentiment accounting for a large fraction of explained variance. Low-payroll-propensity business formation is twice as volatile as high-payroll propensity entrepreneurship, and similarly to noise-based decision making in behavioral finance, it is substantially driven by sentiment.

Keywords: sentiment, entrepreneurship, business formation, push- and pull-motives, behavioral finance

JEL classification: G40, D91, L26, G17

1 Introduction

The determinants of business formation have been studied from a variety of perspectives within the entrepreneurship literature. A large body of work has focused on the distinction between 'opportunity-pull' and 'necessity-push' motivations for entrepreneurship (Schoar, 2010; Ardagna and Lusardi, 2010; Fairlie and Fossen, 2019). Another strand of work focuses on the psychological traits which predispose individuals to entrepreneurship (e.g. Baron, 2000; Frese and Gielnik, 2014). A third strand, international comparative entrepreneurship research, exploits Global Entrepreneurship Monitor data[1] to discover factors underpinning differences between countries' entrepreneurial attitudes, activity, and aspirations, as well as the ways in which these can be influenced by government policies to enhance entrepreneurial activity and economic growth (e.g. Sternberg and Wennekers, 2005; Bergmann et al., 2014).

Yet, the creation of a new business is both an act of entrepreneurship as well as a financial undertaking: finance and entrepreneurship overlap. Here we exploit this overlap. We adapt concepts, tools, and data from behavioral finance to study the determinants of business formation. This approach leads to several layers of novelty.

The data we use to gauge business formation has only recently been made public: the US Census Bureau's weekly business formation statistics (Bayard et al., 2018). Aside from appearing in working papers of the Census Bureau and the Federal Reserve, it has not yet been utilized in the academic entrepreneurship literature. Through the lens of behavioral finance, which distinguishes between fundamentals-based traders and noise-based traders, we distinguish between two different classes of entrepreneurs: those whose information processing is fundamentals-focused, and those whose information processing does not reliably distinguish between fundamentals-based signal and 'noise' from non-fundamental sources. The latter class is susceptible to market sentiment – widespread mood or affect. Following its success as a predictor of market returns (Lemmon and Portniaguina, 2006), we introduce the Michigan Index of Consumer Sentiment ($MICS$) into econometric tests to gauge its power to predict future business formation. And because of the special features of the Census Bureau's business-formation statistics, we are able to estimate the

[1] https://www.gemconsortium.org/

effect of sentiment separately for firms that have (*i*) a high likelihood of supporting a payroll within six months (*HBA*), and (*ii*) a low likelihood of supporting a payroll within six months (*LBA*). These elements are novel to the study of entrepreneurship: the key data and variables, the distinction between fundamentals and noise, and the consequent potential relevance of sentiment.

We find that aggregate business formation is jointly determined by both economic fundamentals and sentiment. Consumer sentiment ($MICS$) predicts month-ahead business formation positively and significantly, accounting for a majority (62.4%) of the explained variance. Other significant predictors of month-ahead business formation are fundamental variables: the composite Purchasing Managers' Index (PMI) and the NBER recession indicator ($RECES$). While $MICS$ and PMI gauge opportunity-pull motives for entrepreneurship, $RECES$ proxies the necessity-push motive for entrepreneurship.

But this picture changes substantially once high-propensity business formation is distinguished from low-propensity business formation. Sentiment is a significant predictor of month-ahead national *low*-propensity business formation, and accounts for 48% of explained variation. However sentiment only accounts for 7.1% of month-ahead national *high*-propensity business formation. This is the first indication that *LBA* and *HBA* entrepreneurs respond differently to information. The second indication is that although the coefficients on the short-term real interest rate variable ($T30R$) are significant in both the *LBA* and *HBA* models, they are nevertheless *of opposite algebraic sign.* For high-propensity entrepreneurs, $T30R$ has a *positive* effect on business formation. In contrast for low-propensity entrepreneurs $T30R$ has a *negative* effect on business formation. The third indication is that whereas PMI ('pull' motive) and $RECES$ ('push' motive) continue to be significant for *LBA*, they are no longer significant for *HBA*. Thus, not only is business formation predicted by both fundamentals and sentiment, but there are two classes of entrepreneurs, differing by their propensity to support a payroll within six months, that respond differently to sentiment and particular fundamentals. Taken together, these results support the interpretation of *HBA* as being fundamentals-oriented entrepreneurs who primarily respond to pull motives, while *LBA* are not exclusively fundamentals-oriented (i.e. also noise- and sentiment-oriented) entrepreneurs who respond to both push and pull motives.

The sequel is organized as follows. Section 2 presents a primer on the terminology and concepts of entrepreneurship used in this paper. Section 3 roots the sentiment hypothesis in a formal model of entrepreneurship. Section 4 presents the data that is used in this paper. Section 5 presents the models and results, including an array of robustness checks. Section 6 concludes.

2 Entrepreneurship

Some of the terminology and concepts used in this paper are not common knowledge among finance specialists.

New business formation is a fundamental feature of entrepreneurship, and it has been viewed as such consistently over time. Joseph Schumpeter defined entrepreneurship as "the assumption of risk and responsibility in designing and implementing a business strategy or starting a business" (Schumpeter, 1911). Subsequently John W. Gough explained that the term entrepreneur "refers to a person who undertakes and operates a new enterprise or venture, and assumes some accountability for the inherent risks" (Gough, 1969). And in the current century, Klapper et. al. (2010) have defined entrepreneurship as "The activities of an individual or a group aimed at initiating economic activities in the formal sector under a legal form of business."

The concepts of 'necessity-push' and 'opportunity-pull' motivation for entrepreneurship were initially shaped by two influential studies. The first, by Gilad and Levine (1986), proceeds within the situational, contingent approach to entrepreneurship which emphasizes external environmental factors over internal psychological traits. They implement empirical tests to determine "which particular environmental factors elicit or hinder the entrepreneurial response" – i.e. business formation. Under their contingent approach, "...people are pushed into entrepreneurship by negative situational factors such as dissatisfaction with existing employment, loss of employment, and career setback." Meanwhile the contingent pull hypothesis hinges upon early experiences (personal, or family), or early training, which encourage the search for profitable business opportunities. And of course attractive external opportunity can also present itself serendipitously, but the metrification of where and when such opportunities arise is not straightforward.

The second influential study, by Amit and Muller (1995), proceeds within the internal-triggers approach to entrepreneurship which emphasizes psychological traits and motives. Accordingly they frame the push hypothesis in terms of an individual's personal dissatisfaction with current employment or lack of ability or motivation to thrive in a current position. And under the internal-trigger framing of the pull hypothesis, an individual is lured by a new venture idea, due to its attractiveness and its favorable personal implications.

The internal-triggers approach to the push and pull hypotheses can be viewed as an extension of the large literature which develops the psychological approach to entrepreneurship (e.g. Baron, 2000; Frese and Gielnik, 2014). On the level of individual psychology and decision making, this literature has shown that the distinguishing feature between entrepreneurs and non-entrepreneurs is not primarily rooted in risk tolerance or risk aversion, but in over- or under-assessment of risk (Licht and Siegel, 2006). Hence it is a question of how individuals process information – which has trait components (trait optimism or pessimism) and transitory components (mood and affect). Behavioral finance also recognizes the role of mood and affect in individual decision making, but focuses on measures which gauge widespread, correlated mood and affect across investors, i.e. *sentiment*. We formalize this connection between business formation and sentiment in Section 3 below.

Contemporary formulations of 'necessity' and 'opportunity' entrepreneurship are aligned with the situational, contingent approach which emphasizes external environmental factors over internal psychological traits and states (Ardagna and Lusardi, 2010; Schoar, 2010; Hurst and Pugsley, 2011; Decker et al., 2014; Fairlie and Fossen, 2019). A wide range of operational definitions are in use. Fairlie and Fossen's (2019) proposal aims to capture consensus, while at the same time remaining consistent with the standard economic model of entrepreneurship:[2] "individuals who are initially unemployed before starting businesses are defined as 'necessity' entrepreneurs, and individuals who are wage/salary workers, enrolled in school or college, or are not actively seeking a job are defined as 'opportunity'

[2]See Evans and Jovanovic (1989).

entrepreneurs." Some authors use different labels for essentially the same distinction, e.g. subsistence or remedial entrepreneurship vs. transformational entrepreneurship (Schoar, 2010; Ardagna and Lusardi, 2010).

In the present paper, both necessity-push and opportunity-pull variables turn out to be empirically important, as do both internal-trigger and external-trigger variants of this distinction.

3 Sentiment hypothesis

The role of sentiment in business formation follows from the standard economic model of entrepreneurship when one of its variables is augmented with a behavioral interpretation. Building upon Fairlie and Fossen (2019), an individual's non-entrepreneurial (outside option) total annual income Y^W is

$$Y^W = w\varepsilon_w + rA \quad , \tag{1}$$

where w is their annual market wage, ε_w is a wage shock, r is the annual interest rate, and A is the individual's assets. If the individual switches to entrepreneurial self-employment, their annual income Y^{SE} is

$$Y^{SE} = \theta f(k)\varepsilon_e + r(A - k) \quad , \tag{2}$$

where θ represents entrepreneurial ability, $f(\cdot)$ is the entrepreneurial production function for annual profits using capital k, and ε_e represents a production shock. The last term in (2) represents the annual interest earned from investing the residual of assets not deployed in the start-up. Entrepreneurial self-employment is chosen when

$$Y^{SE}|_{k=k*} > Y^W \quad . \tag{3}$$

For instance, a downward shock to the wage $\varepsilon_w < 1$ (e.g. partial or full unemployment) may cause (3) to hold, even though it would not be the case in the absence of the wage shock. This is the necessity-push hypothesis. On the other hand, inequality (3) may hold because of an upward shock $\varepsilon_e > 1$ to the production term in (2). This is the opportunity-pull hypothesis. Here the ε_e term aggregates shocks from various sources, including not only contemporaneous demand shocks and production shocks, but also shocks to forward-looking beliefs (expectations) about future demand and production. This is the term that avails of a behavioral interpretation.

An individual whose positive mood skews their processing of information toward perceiving greater entrepreneurial opportunity experiences an upward shock in their personal ε_e term ceteris paribus. The component of mood that is correlated across individuals in the economy – i.e. *sentiment* – is also captured in the ε_e terms across the economy. Positive sentiment shocks – which can be detected and measured at the aggregate level – ceteris paribus increase $Y^{SE}|_{k=k*}$ relative to Y^W, increasing the probability of an individual shifting from wage employment into entrepreneurial self-employment. Conversely negative sentiment shocks ceteris paribus decrease the probability of an individual forming a new business.[3]

To summarize, whereas the external pull and push hypotheses are rooted in standard economic interpretations of ε_w and ε_e, the internal pull hypothesis – i.e. the sentiment hypothesis – follows from a behavioral extension of ε_e.

4 Data

4.1 Business formation

US business formation statistics are available from the US Census Bureau.[4] These statistics are compiled from information disclosed on Form SS-4, the IRS Application for obtaining an Employer Identification Number (EIN).[5] We study national and regional (Northeast, Midwest, South and West respectively) series for Business Applications (*BA*)

[3]A real-options formulation of (3) features an additive, positive-value option term on the right-hand side, which introduces hysteresis in the transition from paid work to entrepreneurial self-employment. Discrete changes in ε_w can influence the magnitude of this option term, and thus the width of the hysteresis band. For instance if an individual loses current employment, $\varepsilon_w = 0$, then the option value is extinguished and the width of the hysteresis band collapses to zero. However, neither the option term itself nor its dependence upon ε_w changes the *signs* of the marginal effects of ε_w and ε_e upon the probability of transition into entrepreneurial self-employment.

[4]https://www.census.gov/econ/bfs/index.html

[5]https://www.irs.gov/forms-pubs/about-form-ss-4

and High-propensity Business Applications (*HBA*). The *BA* series are broad measures of business formation that the Census Bureau characterizes as their 'core business applications series'.[6] The *HBA* series are subsets of the corresponding *BA* series, including only those applications that have a high likelihood of becoming businesses with a payroll.[7] The difference between BA and *HBA* is recorded as Low-propensity Business Applications (*LBA*).

Beginning with the US Census Bureau's weekly frequency, not-seasonally-adjusted data, we first aggregate the series up to monthly frequency, and then remove seasonality from each monthly time series through seasonal-trend decomposition using LOESS (STL) as in Cleveland et al. (1990). These seasonally-adjusted time series are denoted as *BA*-sa, *HBA*-sa and *LBA*-sa, respectively. Figure 1 illustrates the raw, seasonally adjusted, and seasonal components of national *BA*, *HBA* and *LBA*. The sample period spans from 2006M1 to 2018M12. Regional *BA*, *HBA* and *LBA* series have similar seasonal-decomposition patterns. Throughout the analysis in this paper, we focus on seasonally-adjusted measures of business formation, and omit the '-sa' suffix.

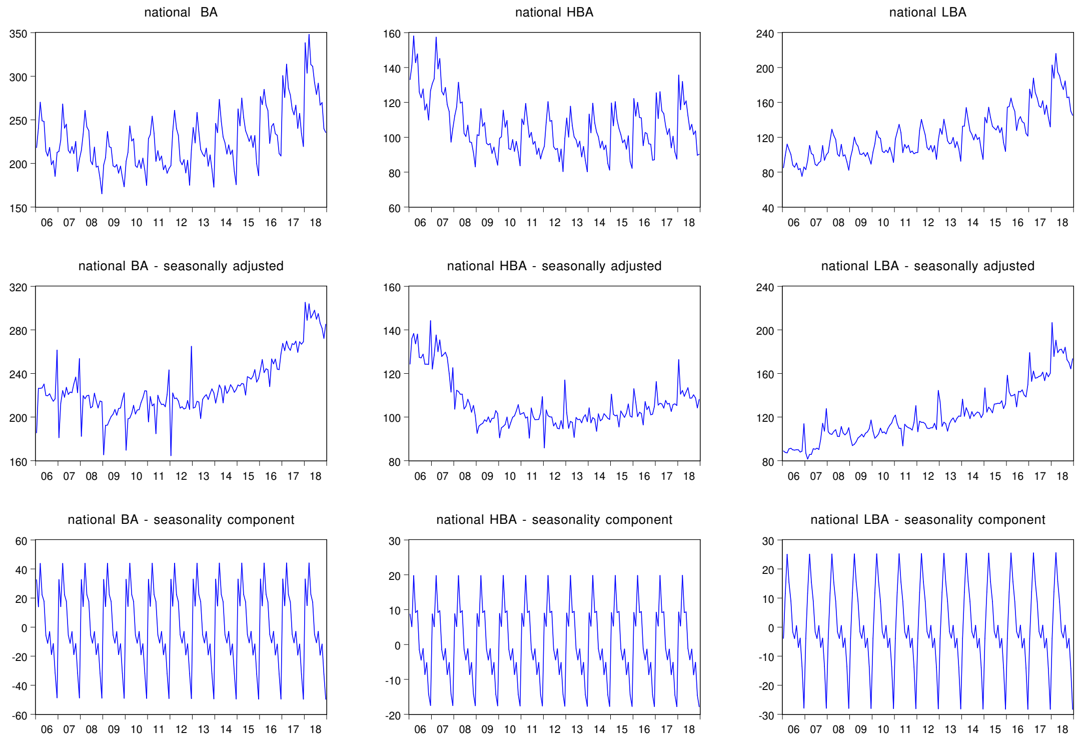
Figure 2 plots the time series of seasonally-adjusted national and regional business formations. For *BA*, *HBA* and *LBA*, the South region accounts for around 41% of the national total, followed by the West region (around 23-24% of the national total). Both the Northeast and Midwest regions each account for 17.5% of the national total. *HBA* fell during the 2007-09 period, as a result of the financial crisis. Thereafter it grew gradually throughout the remainder of the sample period. Meanwhile *LBA* has been growing throughout the entire sample period, without dropping during the 2007-09 period.

Table 1 reports descriptive statistics for the dependent variables. It is noteworthy that the standard deviation of national *LBA* is more than twice that of national *HBA*.[8] National *HBA* is highly right-skewed ($1.36 > 1$) while *LBA* is moderately right-skewed

[6] These series exclude "applications outside of the 50 states and the District of Columbia or those with no state-county geocodes, applications with a NAICS sector code of 11 (agriculture, forestry, fishing and hunting) or 92 (public administration), and applications in certain industries (i.e. private households, certain financial services, civic and social organizations)" (U.S. Census Bureau, 2020).

[7] US Census Bureau defines high-propensity applications as those "(a) for a corporate entity, (b) that indicate they are hiring employees, purchasing a business or changing organizational type, (c) that provide a first wages-paid date (planned wages); or (d) that have a NAICS industry code in manufacturing (31-33), retail stores (44), health care (62), or restaurants/food service (72)" (U.S. Census Bureau, 2020). For the relationship to the Longitudinal Business Database (LBD), see Jarmin and Miranda (2002).

[8] *LBA* sd in the south region is the dominant driver of the national *LBA*'s sd.



(a) *BA* (b) *HBA* (c) *LBA*

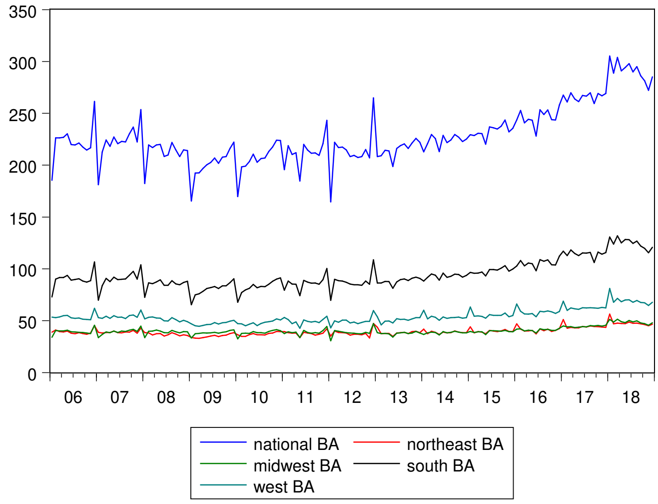
Figure 1: STL decomposition of national *BA*, *HBA* and *LBA*

$(1 > 0.925 > 0.5)$. Similarly, compared to national *LBA*, the national *HBA* distribution has more mass in the tails relative the rest of the distribution (kurtosis $4.22 > 3.34$).

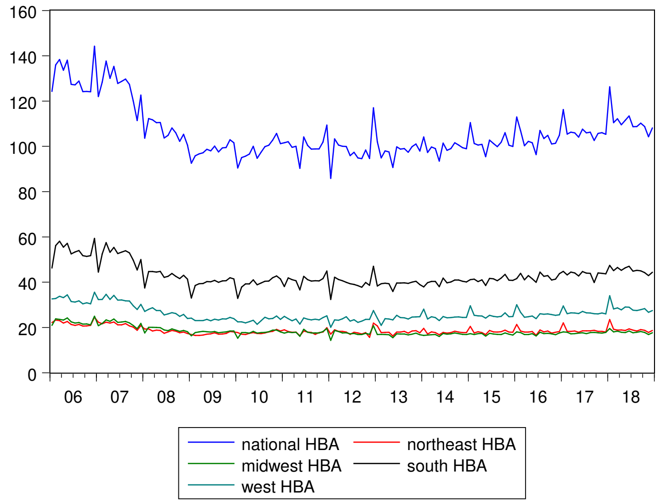
4.2 Sentiment

The most straightforward and direct indicators of sentiment are provided by survey data. Shiller (1999) suggests that the Yale School of Management Stock Market Confidence Indices can reflect the attitudes of institutional investors. In the behavioral asset pricing literature, Qiu and Welch (2006) show that data from the UBS/Gallup surveys can explain equity returns, particularly small-stock returns and returns of stocks held disproportionately by retail investors. Similar findings have also been obtained by Lemmon and Portniaguina (2006) with data from both the Index of Consumer Confidence and the University of Michigan Consumer Confidence Index. Brown and Cliff (2005) find significant long-horizon explanatory power in the Investors Intelligence survey to predict asset prices.

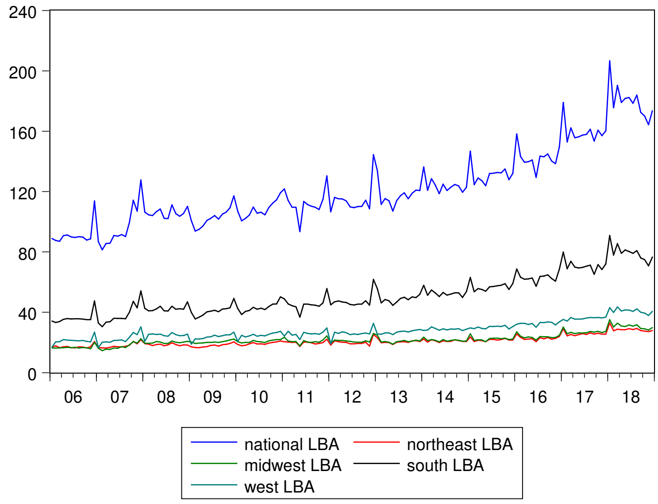
We adopt the monthly Michigan Index of Consumer Sentiment as our proxy for senti-



(a) *BA*



(b) *HBA*



(c) *LBA*

Figure 2: Plots of national and regional series of *BA*, *HBA* and *LBA* (seasonally adjusted)

Table 1: Summary statistics for business formation

	Mean	Median	S.D.	Skewness	Kurtosis
BA national	228.093	221.549	27.493	0.759847	3.565726
BA Northeast	39.538	38.485	3.796	1.210340	5.062385
BA Midwest	40.176	39.356	3.597	0.952617	4.465086
BA South	94.294	90.170	13.746	0.806599	3.234402
BA West	54.085	52.860	6.580	1.155510	4.423357
HBA national	106.169	102.033	11.403	1.359071	4.220803
HBA Northeast	18.660	18.178	1.664	1.283348	4.150596
HBA Midwest	18.341	17.726	1.960	1.434101	4.538684
HBA South	43.061	41.471	5.052	1.247508	4.457796
HBA West	26.105	24.971	3.236	1.030890	3.333198
LBA national	121.924	114.675	25.848	0.924895	3.337371
LBA Northeast	20.877	20.235	3.370	1.048106	3.708815
LBA Midwest	21.835	20.996	3.816	0.970735	4.097016
LBA South	51.232	47.354	13.178	0.828695	2.934472
LBA West	27.980	26.662	5.734	0.748200	3.151906

This table shows summary statistics for seasonally-adjusted national and regional series of *BA*, *HBA* and *LBA*. The full monthly sample contains 156 observations from January 2006 through December 2018.

ment. It is calculated as a linear transformation of the percentages of positive and negative responses on five telephone-survey questions. The five questions cover (i) change in perceived household financial situation over the last year, (ii) expected year-ahead change in household financial situation, (iii) expected year-ahead national financial business conditions, (iv) expected national business conditions (continuous good times vs. periods of widespread unemployment or depression) over the coming 5 years, and (v) current purchasing conditions for major household durable items.

We standardize the indicator and denote the new series as $MICS$.[9] For $MICS$, the sample also covers 2006M1 to 2018M12.

4.3 Fundamental variables

Augmenting the set of fundamental variables employed by behavioral finance sentiment studies to adequately capture the real economy, we assemble fundamental variables from six categories: monetary conditions, real-economy consumption conditions, real-economy production conditions, financial-market conditions, labor-market conditions, and GDP

[9]Here standardization means subtracting the mean and then dividing by the standard deviation.

conditions. This set of fundamental variables combines those that are suggested by the Baker and Wurgler (2007) consumption-based capital asset pricing approach as well as those suggested by the Brown and Cliff (2005) conditional asset pricing approach.

(i) monetary conditions: CPI-based monthly inflation ($INFL$) and 1-month real US Treasury bill return ($T30R$);

(ii) real-economy consumption conditions: real growth rate in total consumption ($CONS$);

(iii) real-economy production conditions: real growth rate in industrial production ($PROD$) and composite Purchasing Manager's Composite Index (PMI);[10]

(iv) financial-market conditions: the spread between 3-month and 1-month real US treasury bill returns ($SPR3$), the spread between 10-year and 3-month real US treasury bill returns ($SPR10$), and the default spread between yields on Moody's Baa- and Aaa-rated corporate bonds ($SPRD$).

(v) labor-market conditions: growth rate in employment ($EMPL$);

(vi) GDP conditions (growth vs. contraction): NBER recession dummy ($RECES$).

$T30R$, $SPR3$, $SPR10$ and $SPRD$ data are obtained from the US Federal Reserve. PMI data are compiled by IHS Markit. Data for the rest fundamental variables are obtained from Jeffrey Wurgler's online data library.

4.4 Sample characteristics

Descriptive statistics for the explanatory variables introduced in Sections 4.2 and 4.3 are summarized in Table 2. With the exception of $MICS$ all variables are more heavy-tailed than the normal distribution. The recession dummy ($RECES$), the 1-month short rate ($T30R$), and three variables representing financial market conditions ($SPR3$, $SPR10$ and $SPRD$) are right-skewed (probably due to the implicit left-truncation for each indicator) while all other indicators are left-skewed.

[10] A PMI value above 50 indicates that purchasing activity has improved month-on-month, whereas a PMI value below 50 indicates that purchasing activity has deteriorated month-on-month.

Table 2: Summary statistics of explanatory variables

	Mean	Median	S.D.	Skewness	Kurtosis
$MICS$	-0.0000	0.0669	1.0013	-0.2876	2.0717
$INFL$	0.0015	0.0017	0.0039	-1.1010	7.6157
$T30R$	0.0009	0.0001	0.0014	1.5667	3.9616
$CONS$	0.0029	0.0033	0.0034	-1.4552	7.8501
$PROD$	0.0006	0.0014	0.0072	-2.1061	12.8744
PMI	52.9962	53.0000	5.0584	-1.5363	6.5693
$SPR3$	0.0001	0.0001	0.0003	4.6122	33.4901
$SPR10$	0.0028	0.0011	0.0190	0.5351	4.8845
$SPRD$	0.0110	0.0094	0.0050	2.6592	10.7802
$EMPL$	0.0007	0.0012	0.0018	-2.0134	7.2590
$RECES$	0.1161	0.0000	0.3214	2.3964	6.7425

This table shows summary statistics for the data of eleven explanatory variables used in the analysis. The full monthly sample contains 156 observations from January 2006 through December 2018.

Table 3 reports pair-wise correlations across the full set of variables. Most correlations (34 out of 55) are statistically significant. $MICS$ is significantly correlated with 8 out of 10 fundamental indicators.

Table 3: Correlation coefficients between indicators

Correlation t-Statistic (Probability)	MICS	INFL	T30R	CONS	PROD	PMI	SPR3	SPR10	SPRD	EMPL	RECES
MICS	1.00										
INFL	0.03	1.00									
	0.373										
	(0.709)										
R30T	0.23	0.13	1.00								
	2.931	1.644									
	(0.004**)	(0.102)									
CONS	0.25	0.51	0.10	1.00							
	3.257	7.253	1.204								
	(0.001**)	(0.000***)	(0.231)								
PROD	0.20	0.10	0.03	0.37	1.00						
	2.519	1.289	0.376	4.887							
	(0.013*)	(0.199)	(0.707)	(0.000***)							
PMI	0.46	0.19	−0.03	0.45	0.55	1.00					
	6.367	2.402	-0.326	6.192	8.166						
	(0.000***)	(0.018*)	(0.745)	(0.000***)	(0.000***)						
SPR3	−0.03	−0.05	0.23	−0.12	−0.15	−0.18	1.00				
	-0.397	-0.670	2.975	-1.536	-1.832	-2.210					
	(0.692)	(0.504)	(0.003**)	(0.127)	(0.069)	(0.029*)					
SPR10	−0.17	−0.28	−0.04	−0.30	−0.03	−0.09	0.23	1.00			
	-2.077	-3.605	-0.479	-3.956	-0.373	-1.058	2.969				
	(0.040*)	(0.000**)	(0.633)	(0.000***)	(0.709)	(0.292)	(0.004**)				
SPRD	−0.58	−0.24	−0.16	−0.48	−0.50	−0.84	0.10	0.11	1.00		
	-8.790	-3.098	-2.088	-6.680	-7.228	-19.077	1.274	1.351			
	(0.000***)	(0.003**)	(0.038*)	(0.000***)	(0.000***)	(0.000***)	(0.205)	(0.179)			
EMPL	0.56	0.16	0.03	0.41	0.47	0.71	−0.15	−0.07	−0.81	1.00	
	8.463	2.001	0.342	5.552	6.657	12.370	-1.909	-0.810	-17.165		
	(0.000***)	(0.047*)	(0.733)	(0.000***)	(0.000***)	(0.000***)	(0.058)	(0.419)	(0.000***)		
RECES	−0.54	−0.00	−0.01	−0.37	−0.56	−0.70	0.27	0.04	0.74	−0.77	1.00
	-7.876	-0.033	-0.071	-4.873	-8.301	-12.011	3.521	0.466	13.532	-14.758	
	(0.000***)	(0.974)	(0.943)	(0.000***)	(0.000***)	(0.000***)	(0.000**)	(0.642)	(0.000***)	(0.000***)	

This table records the correlation coefficients, associated t-statistics and p-values between eleven explanatory variables. *, ** and *** represent significance at 5%, 1% and 0.1% levels, respectively.

5 Models and Results

5.1 Business formation and consumer sentiment

In this section we test for interactions between business formation measures and consumer sentiment, through a Vector Error Correction (VEC) Model. We first confirm that the national and regional business formation measures as well as the consumer sentiment proxy are all persistent time series, following an $I(1)$ processes. Furthermore, Johansen Cointegration tests show that $MICS$ is cointegrated with all fifteen business formation measures (*BA*, *LBA*, and *HBA*, for the national and four regional measures).[11]

Equation (4) presents the VEC model to be estimated, where $Y_t = [MICS_t, BA_t]$ for the *BA* series, $Y_t = [MICS_t, HBA_t]$ for the *HBA* series, and $Y_t = [MICS_t, LBA_t]$ for the *LBA* series. ECT is an error correction term that measures the deviation of Y_t from its long-run cointegration.

$$\Delta Y_t = c + \Theta * ECT_t + \sum_{i=1}^{k} \Phi^{(k)} * \Delta Y_{t-i}^{(k)} + \epsilon_t \tag{4}$$

where

$$ECT_t = BusinessFormation_t + \beta * MICS_t + \alpha \tag{5}$$

We choose optimal lag order k for the VEC models according to the Schwarz Information Criterion (SIC) a.k.a. Bayesian Information Criterion (BIC). According to this criterion, $k = 2$ is optimal for all *BA* and *LBA* measures. For Northeast *HBA* it is $k = 1$ that is optimal, while for national, Midwest, South and West *HBA* it is $k = 3$ that is optimal.

Table 4 reports estimates for the ECT term for different business-formation measures. The t-statistic is recorded in square brackets below each β estimate. The coefficients suggest that in the long-run cointegration, business formation measures are all positively correlated with consumer sentiment (i.e. $MICS$). Moreover, the t-statistics show that such long-run correlation is often statistically significant, with three exceptions: national, Midwest, and South *HBA*.

[11] Model (3) of the Johansen Cointegration test is adopted, i.e. both the cointegration equation and the Vector Autocorrelation model have intercepts but no trend.

Table 4: Long-run cointegration between business formation and consumer sentiment

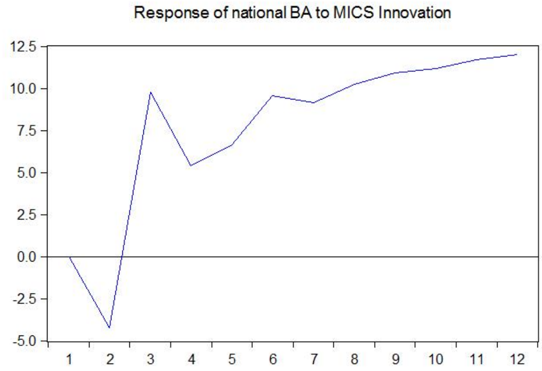
Y_t	β	α
$[MICS_t, nationalBA_t]$	-27.582 [-5.89]	-228.470
$[MICS_t, northeastBA_t]$	-3.168 [-6.132]	-39.538
$[MICS_t, midwestBA_t]$	-3.058 [-3.62]	-40.124
$[MICS_t, southBA_t]$	-14.143 [-5.93]	-94.523
$[MICS_t, westBA_t]$	-6.656 [-5.81]	-54.118
$[MICS_t, nationalLBA_t]$	-27.218 [-3.51]	-122.481
$[MICS_t, northeastLBA_t]$	-3.590 [-3.595]	-20.938
$[MICS_t, midwestLBA_t]$	-3.640 [-2.642]	-21.915
$[MICS_t, southLBA_t]$	-13.819 [-3.77]	-51.526
$[MICS_t, westLBA_t]$	-6.623 [-3.8]	-28.130
$[MICS_t, nationalHBA_t]$	-4.380 [-1.411]	-105.720
$[MICS_t, northeastHBA_t]$	-0.645 [-2.129]	-18.646
$[MICS_t, midwestHBA_t]$	-0.035 [-0.06]	-18.258
$[MICS_t, southHBA_t]$	-2.133 [-1.65]	-42.890
$[MICS_t, westHBA_t]$	-1.909 [-2.30]	-25.998

This table records estimated coefficients for Equation (5). t-statistics are recorded in square brackets below each β estimate.

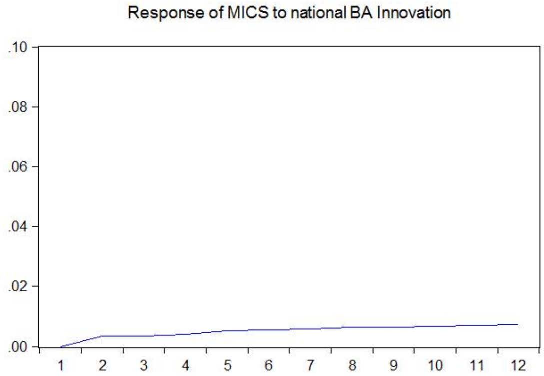
In order to demonstrate the short-run dynamics between business formation and consumer sentiment that is captured by the VEC model, we plot in Figure 3 the impulse response curves between consumer sentiment and national business-formation measures. Figures generated from regional-business formation measures behave similarly and are available upon request.

To generate the impulse response curves, a unit shock is introduced in the impulse variable, and the response period is set at 12 months. For instance, panel (a) of Figure 3 plots the response of national BA to a one-unit shock in $MICS$ over the subsequent

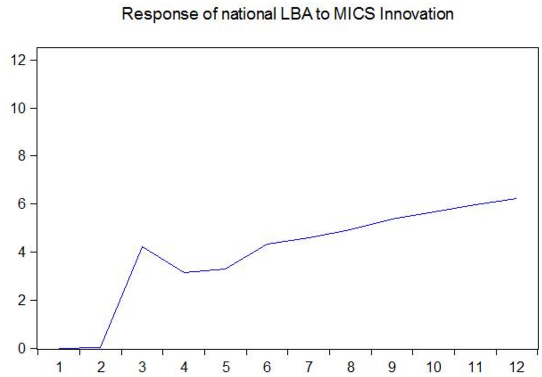
12 months. Business-formation measures are responsive to shocks in consumer sentiment, while consumer sentiment does not respond much to shocks in business formation. Low-propensity business formation and high-propensity business formation respond similarly to shocks in sentiment in the first three months. However from the fourth month onward the effect recedes for *HBA* while further strengthening for *LBA*. This is a first indication — which is consistently reinforced in subsequent empirical analysis — that consumer sentiment impacts upon low-propensity business formation more heavily than upon high-propensity business formation.



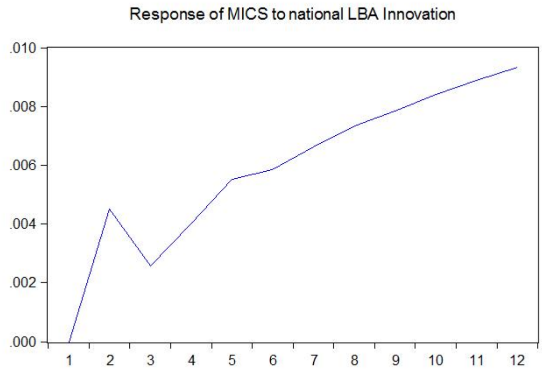
(a) Impulse response, national BA to $MICS$



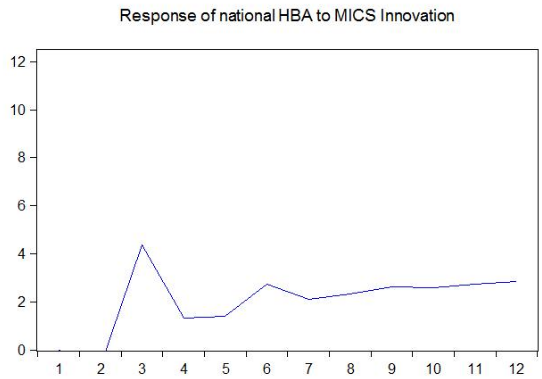
(b) Impulse response, $MICS$ to national BA



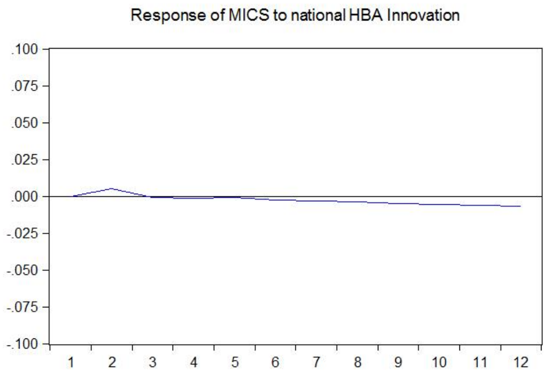
(c) Impulse response, national LBA to $MICS$



(d) Impulse response, $MICS$ to national LBA



(e) Impulse response, national HBA to $MICS$



(f) Impulse response, $MICS$ to national HBA

Figure 3: Impulse-response curves between consumer sentiment and national business formation measures.

5.2 Controlling for fundamentals

In Section 5.1 we find evidence that business formation and consumer sentiment are cointegrated in the long-run, and that such correlation is largely statistically significant. In this section we investigate the incremental explanatory power of consumer sentiment in addition to that of a battery of fundamental variables. We base subsequent estimation on the following multivariate linear equation:

$$BusinessFormation_t = c + \sum \beta^i Factor^i_{t-1} + \epsilon_t \tag{6}$$

Table 5 reports results from estimating Equation (6) with *BA* as the business formation dependent variable. [12] Several findings emerge.

First, the model explains a large proportion of overall business-formation variation. Adjusted R^2 ranges from 0.425 to 0.618. The variance explained is largest for the West region, and somewhat smaller for the Midwest region.

Second, consumer sentiment is a strong predictor of business formation. $MICS$ has positive coefficient estimates in all five *BA*-measure models, and these coefficients are all statistically significant at the 0.1% level. We interpret these result as evidence of a opportunity-pull motivation for entrepreneurship, consistent with the behavioral interpretation of the ε_e term in Section 3. A one standard-deviation increase in $MICS$ on average leads to $22,717$ new businesses being founded monthly in the US, of which more than half $(11,901)$ are located in the South region.

Third, PMI and the NBER recession indicator emerge as the key fundamental variables that explain business formation.

PMI has positive and significant coefficients for national, Northeast and Midwest *BA* series. PMI augments the opportunity-pull motivation driving business formation, where forward-looking improvements in the outlook across manufacturing and services strengthen entrepreneurs' propensity to establish a business, again via its impact upon ε_e. On average a one-unit increase in PMI will lead to an additional $2,191$ monthly business formations across the US.

[12] We verify that residuals from regressing Equation 6 show stationarity, validating the statistical inference.

Table 5: Coefficients of multivariate regressions and p-values: BA.

$Factor^i$	$BusinessFormation$ measures				
	BA national	BA Northeast	BA Midwest	BA South	BA West
$MICS$	22.717	2.632	2.478	11.901	5.621
	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)
$INFL$	-11.849	-133.536	24.158	-37.789	-173.960
	(0.982)	(0.052)	(0.771)	(0.878)	(0.097)
$T30R$	-1604.729	-61.596	10.791	-1463.090	-48.439
	(0.236)	(0.743)	(0.957)	(0.040*)	(0.882)
$CONS$	-239.843	13.443	-47.277	-132.112	-57.211
	(0.699)	(0.841)	(0.604)	(0.668)	(0.652)
$PROD$	325.596	53.152	56.422	152.881	101.775
	(0.200)	(0.086)	(0.090)	(0.201)	(0.030*)
PMI	2.191	0.301	0.466	0.931	0.431
	(0.030*)	(0.031*)	(0.003**)	(0.056)	(0.067)
$SPR3$	5351.224	640.710	1014.622	2138.016	1044.658
	(0.408)	(0.284)	(0.257)	(0.522)	(0.437)
$SPR10$	80.774	-6.314	8.828	38.279	-5.650
	(0.363)	(0.506)	(0.488)	(0.366)	(0.739)
$SPRD$	709.548	119.651	257.126	281.123	80.729
	(0.424)	(0.336)	(0.059)	(0.514)	(0.697)
$EMPL$	-487.110	152.839	-210.155	-251.410	-112.917
	(0.792)	(0.503)	(0.406)	(0.782)	(0.761)
$RECES$	28.884	3.213	3.921	13.315	7.990
	(0.005**)	(0.015*)	(0.004**)	(0.012*)	(0.001**)
$constant$	102.787	21.965	12.291	42.080	29.930
	(0.092)	(0.010**)	(0.196)	(0.153)	(0.035*)
Sample Size	154	154	154	154	154
$adjR^2$	0.578	0.527	0.425	0.612	0.618
relative weight of $MICS$	62.4%	54.1%	50.8%	64.7%	64.0%

This table shows the coefficients and p-values of multivariate regressions in Equation (6), with national and regional BA measures as the dependent variable. p-values are based on Newey-West adjusted standard errors, and lead to no qualitatively different results in hypothesis tests from those based on non-adjusted p-values. *, ** and *** represent significance at 5%, 1% and 0.1% levels, respectively.

The $RECES$ indicator has positive and significant coefficients for national and all regional measures of *BA*, reflecting a clear necessity-push motivation behind business formation during recession periods. On average, 28.9k more businesses are newly founded nationally in each month during recession periods than non-recession periods. The effect is most prominent in the South region, and less so in the Northeast and Midwest regions.

Last but not least, sentiment proxied by $MICS$ is the dominant predictor of *BA* series' variance. More than half of adjusted R^2 can be attributed to consumer sentiment, showing that the broad measure of business formation is heavily responsive to people's sentiment and only partially to entrepreneurs' response to economic conditions. Following the Relative Weight methodology developed by Johnson (2000) and elaborated by Tonidandel and LeBreton (2011), we are able to decompose the adjusted R^2 into proportions attributed to each explanatory variable. The relative importance of $MICS$, as measured by its contribution to adjusted R^2, is reported for each model in the last row of Table 5. For all five *BA* measures, the contribution of $MICS$ exceeds 50%. The relative importance of consumer sentiment is highest in the South, where 64.7% of the explained variance is attributable to $MICS$.

5.3 Decomposition by payroll propensity

Here we investigate high-propensity business applications (*HBA*) separately from low-propensity business applications (*LBA*). We estimate Equation (6) on *HBA* and *LBA* series, separately for the national and each regional level, and report the results in Table 6. It is clear that high-propensity and low-propensity business formation are driven by different sets of factors.

For the *HBA* series, consumer sentiment and the real 30-day t-bill rate stand out as key predictors. Consistent with the findings in Table 5, positive coefficients are present for $MICS$, providing evidence of pull-effect motivation for entrepreneurship. Inflation, manufacturing and services outlook, labour market conditions, and the recession indicator show predictive power inconsistently across the regions.[13]

For the *LBA* series, consumer sentiment, the real 30-day t-bill rate, PMI, and the

[13] Robustness tests confirm that the predictive power found in $INFL$, $PROD$, $EMPL$ and $RECES$ is not robust but rather subject to the choice of standard error adjustments.

NBER recession dummy simultaneously show significant predictive power. Again positive coefficients are present for $MICS$, providing evidence of pull-effect motivation for entrepreneurship. Just as is the case in Table 5, PMI and $RECES$ predict *LBA* with positive coefficients, supporting the interpretation that there is a pull effect from PMI and a push effect from $RECES$.

Interestingly, the coefficient on the 30-day short rate is negative for *LBA* but positive for *HBA*. We interpret this finding as evidence that *HBA* and *LBA* capture different compositions of entrepreneur types: the former comprised of entrepreneurs focusing more on fundamentals, the latter comprised of entrepreneurs who attend to information in a myopic or constrained manner.

In the most basic models of interest-rate determination (e.g. Lucas 1978), an increase in the growth rate of the economy increases the risk-free rate in equilibrium. Non-behavioral fundamentals-focused agents interpret changes in the short-term rate according to fundamentals, and therefore see an increase in the interest rate as a signal of an increase in the growth rate of the economy, leading to an increase in high-propensity business formation (*HBA*). Meanwhile more myopic agents see the short-term rate primarily as a cost increase (which it is, of course), but fail to associate increases in the short-term rate with changes in the economy's growth prospects, and therefore they are less likely to launch a business when short-term borrowing costs increase, leading to a decrease in low-propensity business formation (*LBA*). Hence the positive coefficient on $T30R$ for *HBA* but negative coefficient on $T30R$ for *LBA*.

The difference in the effect of sentiment ($MICS$) among the *HBA* and *LBA* cohorts indeed does support the interpretation that the former are more fundamentals-focused while the latter display non-normative behavioral information processing. But the difference in response to changes in the short-term interest rate could also be plausibly attributed to less business experience and more heavy reliance on short-term borrowing by the *LBA* cohort. This is consistent with the notion that the *HBA* cohort is comprised of more experienced, more well-resourced entrepreneurs who are indeed more likely to support a payroll within six months.

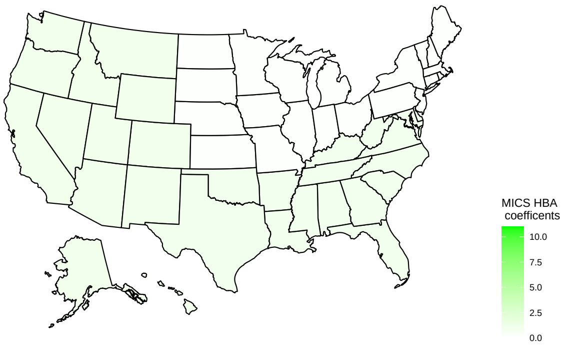
The adjusted R^2 is consistently high across different models reported in Table 6.

However the relative weight of $MICS$ in accounting for the variance explained differs substantially between high- and low-propensity business formation. The *HBA* series are primarily predicted by fundamentals, and the effect of sentiment is much smaller in magnitude than for the *LBA* series (compare Figures 4(a), 4(b)). The $MICS$ coefficients for *LBA* series are 11$\times$ (for national), 19.5$\times$ (for Northeast), 11.4$\times$ (for Midwest), 11.3$\times$ (for South), and 4.4$\times$ (for West) larger than the corresponding *HBA* series. The relative weight of $MICS$ in explaining *HBA* is consistently low and ranges from 1.3% to 14.5% (see Table 6 and Figure 4(c)). In contrast, the *LBA* series are jointly driven by sentiment and economic fundamentals, where the relative weight of $MICS$ ranges from 35.7% to 50.6% (see Table 6 and Figure 4(d)).

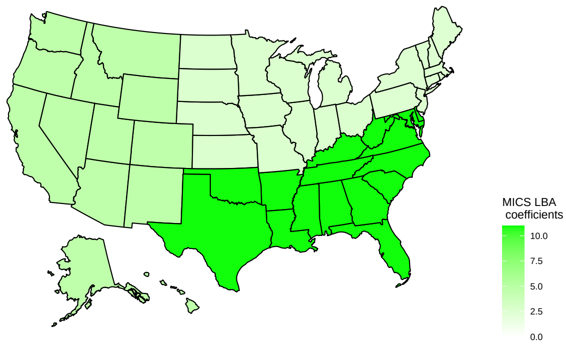
Table 6: Coefficients of multivariate regressions and p-values: *LBA* and *HBA*.

	$BusinessFormation$ measures									
$Factor^i$	*LBA* national	*LBA* Northeast	*LBA* Midwest	*LBA* South	*LBA* West	*HBA* national	*HBA* Northeast	*HBA* Midwest	*HBA* South	*HBA* West
$MICS$	20.751	2.506	2.709	10.916	4.621	1.892	0.128	0.237	0.965	1.054
	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.006**)	(0.279)	(0.038*)	(0.003**)	(0.000***)
$INFL$	-566.567	-82.877	-79.304	-296.796	-21.584	-12.463	-48.061	40.508	86.574	-36.952
	(0.201)	(0.163)	(0.294)	(0.174)	(0.831)	(0.918)	(0.042*)	(0.108)	(0.143)	(0.249)
$T30R$	-8413.977	-930.461	-1229.345	-4403.655	-1855.400	6863.378	867.847	1245.675	2956.309	1787.686
	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)
$CONS$	-147.875	2.405	-38.800	-74.286	-41.285	-85.386	6.663	-8.979	-54.008	-39.412
	(0.769)	(0.968)	(0.626)	(0.773)	(0.717)	(0.521)	(0.719)	(0.742)	(0.504)	(0.228)
$PROD$	286.671	38.427	48.821	126.836	54.970	117.818	14.121	16.678	46.110	36.707
	(0.085)	(0.109)	(0.069)	(0.125)	(0.185)	(0.045*)	(0.231)	(0.089)	(0.197)	(0.031*)
PMI	2.120	0.323	0.458	0.881	0.485	-0.054	-0.217	-0.006	0.018	-0.037
	(0.032*)	(0.015*)	(0.005**)	(0.066)	(0.029*)	(0.812)	(0.623)	(0.888)	(0.852)	(0.478)
$SPR3$	5595.971	744.677	1020.886	2563.669	1501.242	-1171.130	-105.219	-116.520	-582.145	-365.977
	(0.308)	(0.255)	(0.219)	(0.370)	(0.257)	(0.198)	(0.581)	(0.477)	(0.312)	(0.179)
$SPR10$	-3.982	-0.856	-4.019	4.475	11.818	-1.505	-5.928	2.708	12.357	-4.209
	(0.952)	(0.920)	(0.695)	(0.895)	(0.453)	(0.926)	(0.053)	(0.420)	(0.276)	(0.401)
$SPRD$	802.699	121.680	247.886	284.545	133.591	-16.703	-2.263	18.369	18.782	-60.761
	(0.352)	(0.301)	(0.076)	(0.497)	(0.492)	(0.947)	(0.961)	(0.708)	(0.869)	(0.322)
$EMPL$	-568.869	-13.877	-202.897	-240.428	-150.001	217.763	162.524	7.302	25.623	13.106
	(0.689)	(0.944)	(0.330)	(0.741)	(0.652)	(0.621)	(0.020*)	(0.916)	(0.912)	(0.922)
$RECES$	25.891	3.249	3.825	12.380	6.487	2.026	-0.066	-0.021	0.607	1.632
	(0.012*)	(0.015*)	(0.012*)	(0.018*)	(0.004**)	(0.261)	(0.839)	(0.958)	(0.458)	(0.003**)
$constant$	6.559	2.949	-4.215	4.603	1.856	102.978	19.011	17.293	39.215	27.152
	(0.913)	(0.714)	(0.669)	(0.874)	(0.891)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)
$adjR^2$	0.645	0.611	0.576	0.665	0.649	0.784	0.606	0.758	0.745	0.797
relative weight of $MICS$	48.0%	45.7%	37.5%	50.6%	48.2%	7.1%	6.4%	1.3%	7.9%	14.5%

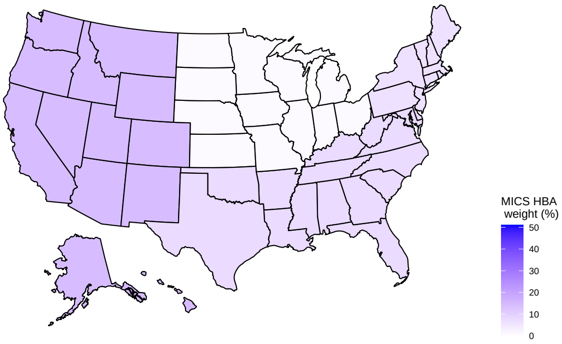
This table shows the coefficients and p-values of multivariate regressions in Equation (6), with national and regional *HBA* and *LBA* measures as the dependent variable. p-values are based on Newey-West adjusted standard errors, and lead to no qualitatively different results in hypothesis tests from those based on non-adjusted p-values. *, ** and *** represent significance at 5%, 1% and 0.1% levels, respectively.



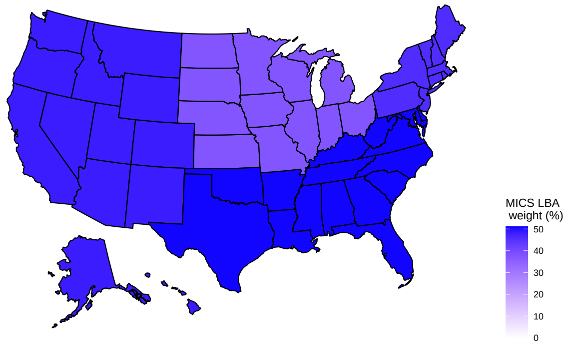
(a) *MICS HBA* coefficients



(b) *MICS LBA* coefficients



(c) *MICS HBA* weight (%)



(d) *MICS LBA* weight (%)

Figure 4: *MICS* coefficients (top) and weights (bottom), separately for *HBA* (left) and *LBA* (right).

5.4 Additional checks

5.4.1 Out-of-sample performance

We assess the out-of-sample (OOS) performance of three competing model specifications using two training samples, two estimation approaches, and three performance criteria. We run these analyses separately for each combination of business-formation outcome (*BA*, *LBA*, *HBA*) and region (National, Northeast, Midwest, South, West).

The three model specifications included in this OOS horse race are: the Grand Mean model (GM), the Fundamentals-only model (F), and the Fundamentals + Sentiment model (F + S).

To ensure results do not hinge on a particular choice of within-sample estimation period, we work with two different training samples: a 10-year sample running from 2006M1 to 2015M12, and an 11-year sample running from 2006M1 to 2016M12. For each training sample, we estimate OOS performance both by fixing the estimation window to the training sample as well as by incrementally extending the estimation window by one month and then recalibrating. In the latter 'rolling-and-recalibrating' forecasting approach, each one-month-ahead forecast benefits from regression coefficients estimated on all previous data points. As performance criteria we report Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

Several strong, consistent findings emerge from the OOS forecasting performance reported in Figures 5–8. First, for *BA* and *LBA*, the Fundamentals-only model improves OOS performance relative to the Mean-only model. Adding Sentiment to Fundamentals greatly improves OOS performance. This performance ranking is consistent across the RMSE, MAE, and MAPE criteria. Second, for *HBA*, differences between the OOS performance of the three different models are much less pronounced; neither Sentiment nor Fundamentals improve OOS performance relative to the Mean-only model. This finding is consistent across training samples, estimation approaches, and geographical regions on both RMSE and MAE performance criteria. On the MAPE performance criterion differences between models are somewhat more pronounced,[14] though the ranking of models

[14] MAPE treats over- and under-predictions asymmetrically, and is known to overstate prediction errors relative to other performance criteria.

varies among training samples, estimation approaches, and geographical regions.

These two findings are consistent with the conclusions of the analysis conducted in Section 5.3: the non-Sentiment and Sentiment models have the same forecasting performance for *HBA* because these entrepreneurs do not confuse the 'noise' of sentiment for fundamental signal. Also note that the standard deviation of *LBA* is more than twice that of *HBA* – there is much more variation to be explained in the former than in the latter. In Table 6, the constant term and $T30R$[15] are highly significant, large-coefficient predictors of *HBA*. The OOS findings reinforce an interpretation whereby the 'innovation process' that generates *HBA*s is relatively stable and not responsive to Sentiment.

[15]Note that following the financial crisis, $T30R$ remained close to zero with very low volatility throughout the sample period, including the OOS period.

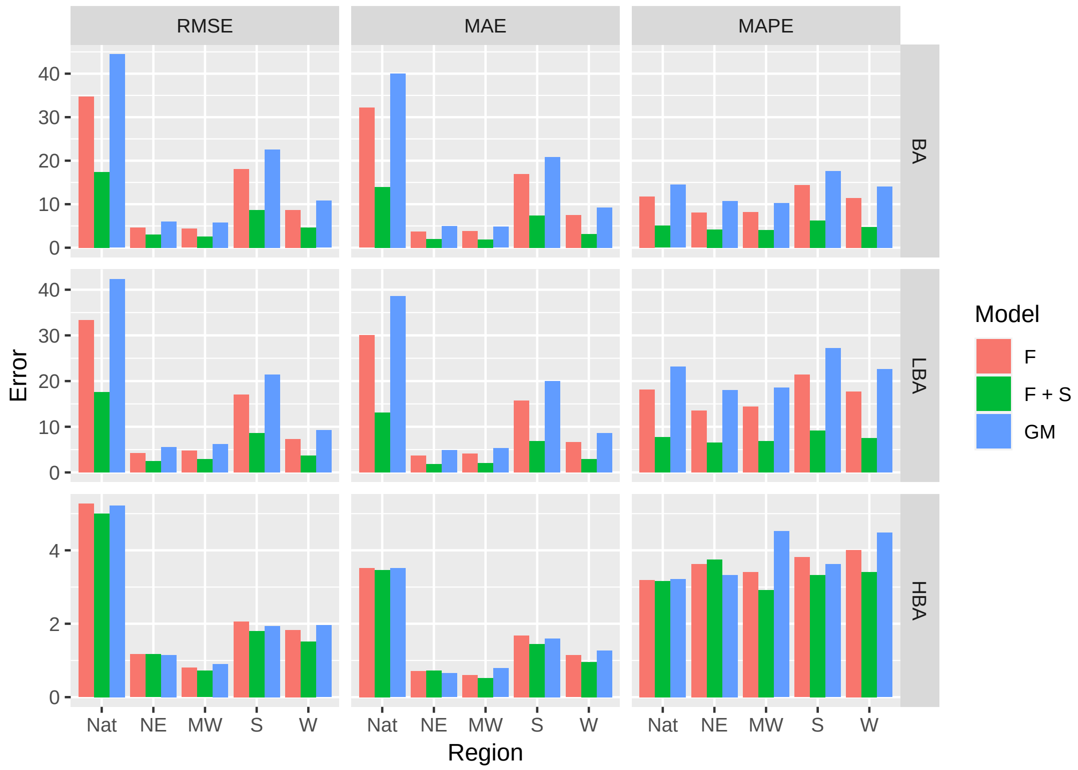


Figure 5: Out-of-sample forecasting performance: fixed 10-year training sample

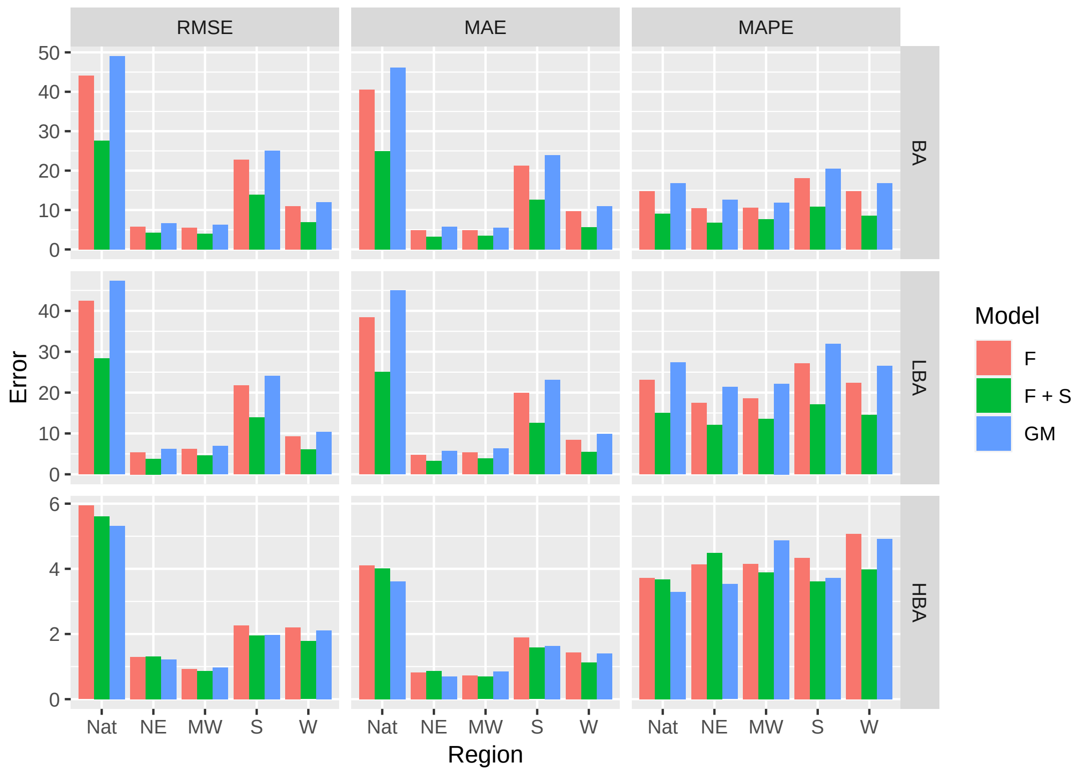


Figure 6: Out-of-sample forecasting performance: rolling-and-recalibrating from a 10-year training sample

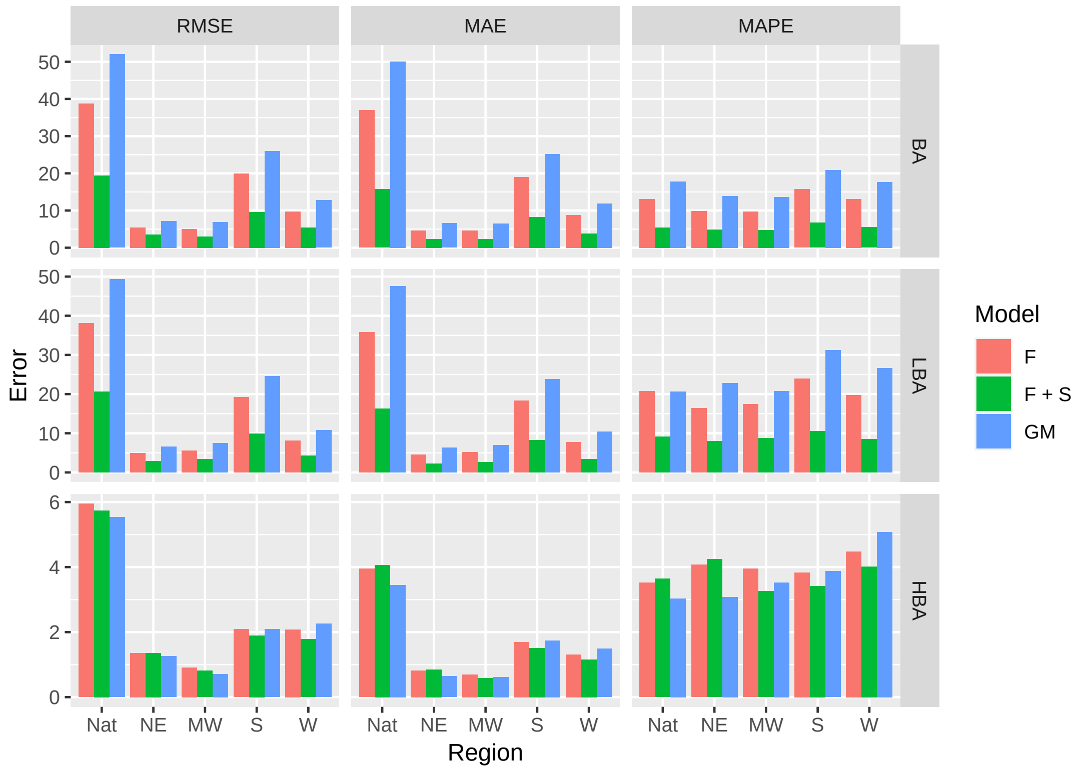


Figure 7: Out-of-sample forecasting performance: fixed 11-year training sample

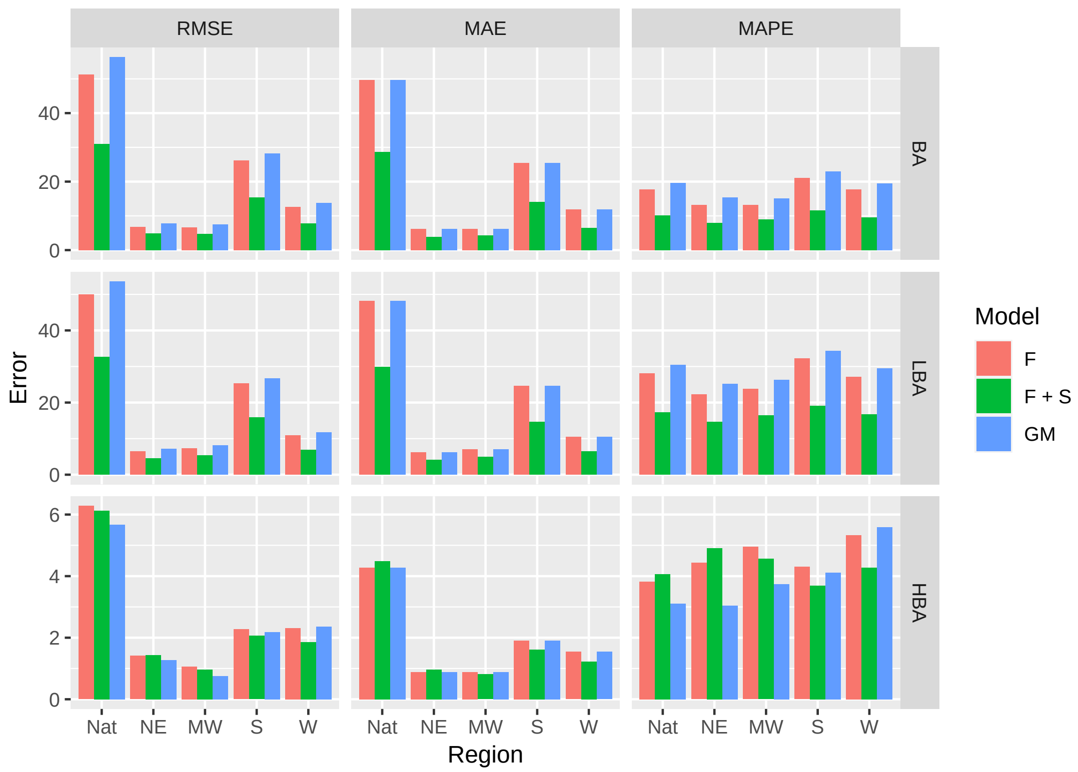


Figure 8: Out-of-sample forecasting performance: rolling-and-recalibrating from an 11-year training sample

5.4.2 Multicollinearity

Many of the variables used in this analysis are correlated — some of them highly and statistically significantly so (see Table 3). These correlations may create multicollinearity problems when estimating Equation (6). Nevertheless multicollinearity between explanatory variables is possible even without high bivariate correlation (Salmerón et al., 2018). This section employs Variance Inflation Factors (*VIF*s) — and in particular the *VIF* > 10 rule of thumb — to identify multicollinearity.[16] We apply a procedure whereby the predictor with the highest *VIF* above 10 is dropped, and the regression is re-estimated. This procedure is implemented recursively until *VIF* values fall below 10 for all remaining predictors. We find that removing $SPRD$ from the bank of predictors will eliminate multicollinearity for all the business-formation measures in Tables 5 and 6 except national *HBA* and Northeast *HBA*. For these two regressions both $SPRD$ and $RECES$ should be removed. The results from the new regressions after eliminating multicollinearity are reported in Tables 7 and 8. These results show that the findings from Tables 5 and 6 remain robust after controlling for multicollinearity.

[16]Different levels of stringency are reflected in the choice of *VIF* cutoff. The cutoff of 10 is used widely, although some researchers advocate 5, or even 2.5.

Table 7: Coefficients of multivariate regressions and p-values, controlling for multicollinearity: *BA*.

$Factor^i$	$BusinessFormation$ measures				
	BA national	BA Northeast	BA Midwest	BA South	BA West
$MICS$	22.435	2.584	2.375	11.789	5.589
	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)
$INFL$	75.781	-144.317	0.991	-63.119	-181.234
	(0.880)	(0.028*)	(0.990)	(0.794)	(0.078)
$T30R$	-1950.793	-119.953	-114.616	-1600.201	-87.813
	(0.110)	(0.481)	(0.523)	(0.014*)	(0.779)
$CONS$	-252.018	11.390	-51.689	-136.936	-58.597
	(0.691)	(0.870)	(0.596)	(0.662)	(0.649)
$PROD$	330.584	53.993	58.230	154.858	102.342
	(0.201)	(0.080)	(0.091)	(0.201)	(0.030*)
PMI	1.848	0.243	0.342	0.795	0.391
	(0.010*)	(0.017*)	(0.003**)	(0.023*)	(0.023*)
$SPR3$	4871.077	559.742	840.627	1947.781	990.029
	(0.437)	(0.333)	(0.311)	(0.550)	(0.453)
$SPR10$	81.907	-6.123	9.238	38.728	-5.521
	(0.347)	(0.513)	(0.455)	(0.353)	(0.744)
$SPRD$					
$EMPL$	-1126.609	45.000	-441.897	-504.780	-185.676
	(0.536)	(0.851)	(0.108)	(0.572)	(0.600)
$RECES$	30.233	3.440	4.410	13.850	8.143
	(0.005**)	(0.011*)	(0.003**)	(0.011*)	(0.001**)
constant	129.526	26.474	21.980	52.674	32.972
	(0.000***)	(0.000***)	(0.000***)	(0.003**)	(0.000***)
$adjR^2$	0.578	0.526	0.407	0.613	0.620
relative weight of $MICS$	65.73%	57.27%	54.4%	67.9%	67.38%

This table shows the coefficients and p-values of multivariate regressions in Equation (6), with national and regional *BA* measures as the dependent variable. Predictors with $VIF > 10$ are removed from the equation to control for multicollinearity. p-values are based on Newey-West adjusted standard errors, and lead to no qualitatively different results in hypothesis tests from those based on non-adjusted p-values. *, ** and *** represent significance at 5%, 1% and 0.1% levels, respectively.

Table 8: Coefficients of multivariate regressions and p-values, controlling for multicollinearity: *LBA* and *HBA*.

	$BusinessFormation$ measures									
$Factor^i$	*LBA* national	*LBA* Northeast	*LBA* Midwest	*LBA* South	*LBA* West	*HBA* national	*HBA* Northeast	*HBA* Midwest	*HBA* South	*HBA* West
$MICS$	20.431	2.458	2.610	10.803	4.568	1.795	0.133	-0.245	0.957	1.078
	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.003**)	(0.205)	(0.025*)	(0.003**)	(0.000***)
$INFL$	-638.892	-93.841	-101.639	-322.434	-33.621	14.841	-48.762	38.853	84.881	-31.478
	(0.154)	(0.115)	(0.180)	(0.139)	(0.742)	(0.897)	(0.025*)	(0.115)	(0.138)	(0.323)
$T30R$	-8805.474	-989.807	-1350.245	-4542.435	-1920.556	6866.117	869.140	1236.717	2947.149	1817.321
	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)
$CONS$	-161.649	0.317	-43.053	-79.169	-43.577	-97.398	7.133	-9.294	-54.331	-38.370
	(0.755)	(0.996)	(0.612)	(0.763)	(0.708)	(0.453)	(0.692)	(0.734)	(0.501)	(0.232)
$PROD$	292.314	39.283	50.564	128.836	55.909	101.317	14.680	16.808	46.242	36.280
	(0.082)	(0.107)	(0.058)	(0.122)	(0.188)	(0.064)	(0.165)	(0.083)	(0.194)	(0.032*)
PMI	1.732	39.283	0.338	0.743	0.420	-0.071	-0.020	-0.015	0.009	-0.008
	(0.014*)	(0.005**)	(0.004**)	(0.029*)	(0.006**)	(0.586)	(0.462)	(0.540)	(0.884)	(0.844)
$SPR3$	5052.790	662.337	853.143	2371.120	1410.841	-825.676	-115.404	-128.950	-594.854	-324.860
	(0.343)	(0.299)	(0.278)	(0.397)	(0.276)	(0.361)	(0.532)	(0.458)	(0.315)	(0.254)
$SPR10$	-2.700	-0.661	-3.624	4.929	12.031	-3.395	-5.867	2.738	12.387	-4.306
	(0.967)	(0.936)	(0.716)	(0.883)	(0.435)	(0.829)	(0.050)	(0.419)	(0.274)	(0.381)
$SPRD$										
$EMPL$	-1292.322	-123.545	-426.311	-496.882	-270.404	77.635	170.004	-9.254	8.695	67.868
	(0.384)	(0.557)	(0.082)	(0.500)	(0.417)	(0.844)	(0.014*)	(0.874)	(0.964)	(0.516)
$RECES$	27.418	3.481	4.297	12.922	6.741			0.014	0.643	1.516
	(0.010*)	(0.013*)	(0.008**)	(0.017*)	(0.005**)			(0.974)	(0.463)	(0.016*)
$constant$	36.808	7.535	5.126	15.326	6.890	103.990	18.869	17.985	39.923	24.863
	(0.300)	(0.113)	(0.392)	(0.374)	(0.367)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)
$adjR^2$	0.644	0.608	0.561	0.666	0.649	0.786	0.611	0.759	0.747	0.797
relative weight of $MICS$	48.96%	47.25%	37.97%	51.54%	49.38%	7.93%	7.73%	1.57%	8.58%	15.56%

This table shows the coefficients and p-values of multivariate regressions in Equation (6), with national and regional *HBA* and *LBA* measures as the dependent variable. Predictors with $VIF > 10$ are removed from the equation to control for multicollinearity. p-values are based on Newey-West adjusted standard errors, and lead to no qualitatively different results in hypothesis tests from those based on non-adjusted p-values. *, ** and *** represent significance at 5%, 1% and 0.1% levels, respectively.

5.4.3 Variable transforms

In this section we discuss two robustness checks: (i) replacing the recession indicator with a Global Financial Crisis (GFC) indicator, and (ii) substituting variables with their log transforms.

The NBER recession indicator used throughout the main analysis sections is a dummy variable for the 01/2008 – 07/2009 period. The spread between LIBOR and the Overnight Indexed Swap rate (OIS) is a measure of credit risk in the banking sector. The US Dollar 3-month LIBOR-OIS interest-rate spread spiked on 9 August 2007, and only returned to pre-crisis levels on 7 May 2009. Hence we define the GFC period in the credit crunch variable $CRECRU$ as 08/2007 – 05/2009.

Results for the GFC dummy $CRECRU$ are reported in columns 2–6 of Tables 9–11. Like the $RECES$ indicator in the *BA* and *LBA* models, $CRECRU$ is significant at the national level as well as in all sub-national regions. In these models, $MICS$ remains significant across all regions at the 0.1% level, but with marginally higher coefficient estimates than with $RECES$. In the *BA* models with $CRECRU$, $T30R$ is significant both at the national level as well as in the South, whereas with $RECES$ it is only significant in the South. Meanwhile in the *LBA* models, $T30R$ has large coefficients significant at the 0.1% level in all regions with both $CRECRU$ and $RECES$. For the *BA* models, PMI is significant at the national level and in all four sub-regions, compared with only two out of four with $RECES$. For the *LBA* models, PMI is significant at the national level and in three out of four sub-regions – the same proportion as with $RECES$. For the *BA* and *LBA* models, substitution of $RECES$ with $CRECRU$ marginally increases the adjusted R^2. Overall the *BA*- and *LBA*-model results are robust to substitution of $RECES$ with $CRECRU$.

For the *HBA* models, two of the regional coefficients for $CRECRU$ are significant, compared with only one for $RECES$. With respect to $MICS$, $T30R$ and the constant term as predictors of *HBA*, there is no difference between the $RECES$ and $CRECRU$ variants. The *HBA*-model results are robust to substitution of $RECES$ with $CRECRU$.

Robustness is explored further by applying logarithmic transforms to $MICS$, PMI, and all regional variants of BA, LBA, and HBA. Results of re-estimating Equation (5) with these log transforms are reported in columns 7–11 of Tables 9–11.

For the BA models, the logarithmic transforms have no effect on the significance of $MICS$. However whereas PMI is a significant predictor of BA nationally, for the Northeast and for the Midwest, $log(PMI)$ is a significant predictor of $log(BA)$ only for the Midwest. The constant term is significant in all of the log-transformed BA models, wheres without log transforms, the constant term is significant only for the Northeast and the South. The adjusted R^2 and the relative weight of $log(MICS)$ is marginally lower than in the models without log transforms.

For the LBA models, only two coefficients' significance is changed by log transforms, and all of the $log(MICS)$ coefficients remain significant at the 0.1% level. The coefficient on $log(PMI)$ becomes non-significant for national and West-region $log(LBA)$. The log transforms reduce the adjusted R^2 marginally from 0.576–0.665 to 0.556–0.646, whereas the relative weight of $log(MICS)$ is reduced by approximately 10%, from 37.5%–50.6% to 28.5%–40.7%.

For the HBA models, only two coefficients' significance is changed by log transforms, and the constant terms remain significant at the 0.1% level across all regions. The coefficient on $log(MICS)$ in the Midwest $log(HBA)$ model drops into non-significance. Outside of any other broader pattern in the estimates, $SPR10$ pops up as significant in the Northeast $log(HBA)$ model. Compared with the models in Table 6, the adjusted R^2 figures are comparable, while the relative weight of $log(MICS)$ is marginally greater than the relative weight of $MICS$.

None of the log-transform results undermine the main results and their interpretations set out in Sections 5.2 and 5.3.

Table 9: Coefficients of multivariate regressions and p-values with robustness checks: BA and $log(BA)$.

$Factor^i$	$BusinessFormation$ measures									
	BA national	BA Northeast	BA Midwest	BA South	BA West	$log(BA)$ national	$log(BA)$ Northeast	$log(BA)$ Midwest	$log(BA)$ South	$log(BA)$ West
$MICS$	23.519	2.753	2.594	12.327	5.878					
	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)					
$log(MICS)$						0.607517	0.405728	0.380005	0.760468	0.634066
						(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)
$INFL$	238.574	-107.922	57.591	73.623	-107.206	-0.220923	-3.513555	0.497328	-0.734457	-3.612371
	(0.620)	(0.093)	(0.463)	(0.748)	(0.275)	(0.921)	(0.034*)	(0.805)	(0.777)	(0.059)
$T30R$	-2543.404	-181.860	-120.576	-1924.320	-325.893	-7.415001	-1.384776	-0.259653	-15.71010	-0.767341
	(0.032*)	(0.247)	(0.485)	(0.003**)	(0.190)	(0.206)	(0.768)	(0.958)	(0.035*)	(0.898)
$CONS$	-348.937	2.173	-61.869	-180.849	-86.418	-0.927705	0.429795	-1.101770	-1.220883	-0.754966
	(0.565)	(0.973)	(0.479)	(0.548)	(0.481)	(0.739)	(0.802)	(0.631)	(0.717)	(0.744)
$PROD$	272.471	49.957	49.888	133.273	90.128	1.391077	1.276815	1.371382	1.533812	1.819479
	(0.266)	(0.094)	(0.121)	(0.244)	(0.044*)	(0.277)	(0.111)	(0.127)	(0.299)	(0.053)
PMI	2.423	0.333	0.500	1.048	0.501					
	(0.020*)	(0.020*)	(0.002**)	(0.036*)	(0.037*)					
$log(PMI)$						0.415184	0.344652	0.551026	0.406113	0.302044
						(0.069)	(0.055)	(0.006**)	(0.127)	(0.176)
$SPR3$	2224.120	171.121	559.791	477.495	42.932	20.24157	13.37342	22.17966	19.01070	14.57205
	(0.652)	(0.696)	(0.398)	(0.851)	(0.964)	(0.482)	(0.366)	(0.317)	(0.599)	(0.551)
$SPR10$	82.035	-5.751	9.105	39.621	-4.826	0.403320	-0.136041	0.255143	0.473979	-0.068610
	(0.365)	(0.555)	(0.483)	(0.359)	(0.781)	(0.329)	(0.571)	(0.438)	(0.315)	(0.828)
$SPRD$	10.840	1.610	3.079	4.531	1.839	0.022592	0.027589	0.064626	0.017440	0.001304
	(0.251)	(0.210)	(0.030*)	(0.317)	(0.408)	(0.597)	(0.426)	(0.081)	(0.728)	(0.976)
$EMPL$	-1046.400	109.730	-281.307	-474.888	-246.177	-1.628756	4.195408	-5.972690	-1.226101	-1.892825
	(0.560)	(0.608)	(0.260)	(0.584)	(0.497)	(0.858)	(0.486)	(0.387)	(0.910)	(0.795)
$RECES$						0.131570	0.081503	0.099734	0.145813	0.150753
						(0.007**)	(0.019*)	(0.005**)	(0.015*)	(0.001**)
$CRECRU$	24.924	3.147	3.476	12.163	7.315					
	(0.004**)	(0.001**)	(0.003**)	(0.004**)	(0.000***)					
$constant$	87.727	19.940	10.159	34.509	25.372	7.985301	7.393980	6.665371	6.476446	6.899180
	(0.166)	(0.023*)	(0.301)	(0.256)	(0.082)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)
Sample Size	154	154	154	154	154	154	154	154	154	154
$adjR^2$	0.580	0.535	0.429	0.617	0.627	0.526	0.516	0.366	0.562	0.592
relative weight of $MICS$	62.87%	54.52%	51.30%	64.79%	64.54%					
relative weight of $log(MICS)$						60.19%	51.38%	49.80%	61.96%	62.08%

This table shows the coefficients and p-values of multivariate regressions in Equation (6). Columns 2 to 5 report results from a regression model same as that in Table 5, except for replacing $RECES$ with the newly defined credit crunch indicator ($CRECRU$). Columns 6 to 10 record results from a regression model that takes logged national and regional BA measures as the dependent variable, as well as logged $MICS$ and logged PMI as replacements for the original series. p-values are based on Newey-West adjusted standard errors, and lead to no qualitatively different results in hypothesis tests from those based on non-adjusted p-values. *, ** and *** represent significance at 5%, 1% and 0.1% levels, respectively.

Table 10: Coefficients of multivariate regressions and p-values with robustness checks: LBA and $log(LBA)$.

$Factor^i$	$BusinessFormation$ measures									
	LBA national	LBA Northeast	LBA Midwest	LBA South	LBA West	$log(LBA)$ national	$log(LBA)$ Northeast	$log(LBA)$ Midwest	$log(LBA)$ South	$log(LBA)$ West
$MICS$	22.175	2.692	2.933	11.608	4.904					
	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)					
$log(MICS)$						0.983	0.702	0.721	1.217	0.962
						(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)
$INFL$	-392.348	-61.528	-54.575	-214.268	27.358	-5.466	-4.325	-4.213	-7.072	-1.072
	(0.327)	(0.253)	(0.423)	(0.281)	(0.765)	(0.118)	(0.109)	(0.205)	(0.085)	(0.759)
$T30R$	-9611.032	-1084.323	-1413.342	-4981.587	-2117.861	-77.500	-48.646	-64.701	-97.200	-74.848
	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)
$CONS$	-226.276	-7.235	-49.993	-111.475	-62.970	-0.585	0.372	-1.230	-0.658	-1.406
	(0.631)	(0.896)	(0.498)	(0.647)	(0.561)	(0.885)	(0.899)	(0.730)	(0.895)	(0.735)
$PROD$	299.933	40.715	52.004	134.122	51.881	2.072	1.651	2.035	2.067	1.746
	(0.056)	(0.068)	(0.044*)	(0.086)	(0.178)	(0.142)	(0.168)	(0.088)	(0.223)	(0.309)
PMI	2.461	0.367	0.511	1.046	0.556					
	(0.012*)	(0.005**)	(0.002**)	(0.027*)	(0.013)					
$log(PMI)$						0.682	0.652	0.916	0.623	0.697
						(0.099)	(0.044*)	(0.016*)	(0.194)	(0.085)
$SPR3$	61.754	22.216	148.388	-125.089	402.301	45.688	33.892	46.538	51.416	55.858
	(0.986)	(0.957)	(0.783)	(0.948)	(0.652)	(0.308)	(0.270)	(0.221)	(0.363)	(0.246)
$SPR10$	6.641	0.575	-2.259	9.702	13.480	0.053	0.014	-0.107	0.210	0.485
	(0.923)	(0.949)	(0.834)	(0.782)	(0.405)	(0.919)	(0.972)	(0.812)	(0.738)	(0.389)
$SPRD$	11.308	1.628	2.962	4.413	2.166	0.041	0.045	0.101	0.020	0.020
	(0.204)	(0.179)	(0.038*)	(0.306)	(0.291)	(0.597)	(0.464)	(0.148)	(0.826)	(0.792)
$EMPL$	-641.797	-18.645	-205.062	-268.653	-213.390	-3.368	0.249	-10.172	-1.830	-4.458
	(0.618)	(0.916)	(0.289)	(0.681)	(0.489)	(0.780)	(0.980)	(0.306)	(0.902)	(0.729)
$RECES$						0.207	0.150	0.172	0.234	0.233
						(0.019*)	(0.025*)	(0.019*)	(0.030*)	(0.007**)
$CRECRU$	30.737	3.943	4.710	14.828	6.817					
	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)					
$constant$	-14.778	0.192	-7.525	-5.722	-2.665	-2.230	-2.663	-3.788	-3.863	-3.654
	(0.804)	(0.981)	(0.444)	(0.842)	(0.845)	(0.325)	(0.134)	(0.070)	(0.139)	(0.100)
Sample Size	154	154	154	154	154	154	154	154	154	154
$adjR^2$	0.677	0.643	0.613	0.695	0.672	0.626	0.591	0.556	0.646	0.615
relative weight of $MICS$	46.45%	44.60%	35.97%	48.95%	46.82%					
relative weight of $log(MICS)$						39.13%	39.59%	28.48%	40.73%	39.54%

This table shows the coefficients and p-values of multivariate regressions in Equation (6). Columns 2 to 5 report results from a regression model same as that in Table 7, except for replacing $RECES$ with the newly defined credit crunch indicator ($CRECRU$). Columns 6 to 10 record results from a regression model that takes logged national and regional LBA measures as the dependent variable, as well as logged $MICS$ and logged PMI as replacements for the original series. p-values are based on Newey-West adjusted standard errors, and lead to no qualitatively different results in hypothesis tests from those based on non-adjusted p-values. *, ** and *** represent significance at 5%, 1% and 0.1% levels, respectively.

Table 11: Coefficients of multivariate regressions and p-values with robustness checks: HBA and $log(HBA)$.

$Factor^i$	$BusinessFormation$ measures									
	HBA national	HBA Northeast	HBA Midwest	HBA South	HBA West	$log(HBA)$ national	$log(HBA)$ Northeast	$log(HBA)$ Midwest	$log(HBA)$ South	$log(HBA)$ West
$MICS$	1.506	0.063	0.318	0.762	0.993					
	(0.020*)	(0.589)	(0.002**)	(0.009**)	(0.000***)					
$log(MICS)$						0.120	0.045	0.078	0.153	0.268
						(006**)	(0.263)	(0.051)	(0.005**)	(0.000***)
$INFL$	36.583	-44.142	45.996	107.431	-15.300	-0.372	-2.589	2.084	1.842	-1.797
	(0.769)	(0.048*)	(0.070)	(0.077)	(0.664)	(0.730)	(0.029*)	(0.118)	(0.173)	(0.120)
$T30R$	7020.307	901.741	1286.501	3046.946	1787.779	60.061	44.219	63.335	63.088	63.323
	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)
$CONS$	-105.184	5.179	-11.087	-62.317	-48.471	-0.587	0.403	-0.295	-0.974	-1.251
	(0.430)	(0.778)	(0.687)	(0.448)	(0.144)	(0.617)	(0.673)	(0.837)	(0.581)	(0.302)
$PROD$	75.956	8.805	9.846	26.101	24.613	1.102	0.705	0.873	1.117	1.368
	(0.202)	(0.416)	(0.306)	(0.483)	(0.146)	(0.055)	(0.250)	(0.104)	(0.210)	(0.047*)
PMI	-0.121	-0.034	-0.021	-0.018	-0.044					
	(0.573)	(0.428)	(0.598)	(0.837)	(0.400)	(0.)	(0.)	(0.)	(0.)	(0.)
$log(PMI)$						-0.05	-0.035	0.019	0.052	-0.077
	(0.)	(0.)	(0.)	(0.)	(0.)	(0.965)	(0.764)	(0.876)	(0.643)	(0.443)
$SPR3$	320.281	145.795	194.005	199.649	-134.712	-8.203	-2.968	-3.233	-9.607	-12.040
	(0.825)	(0.462)	(0.428)	(0.801)	(0.740)	(0.315)	(0.748)	(0.693)	(0.499)	(0.267)
$SPR10$	-7.362	-6.779	1.636	9.438	-5.555	-0.024	-0.325	0.137	0.288	-0.155
	(0.647)	(0.021*)	(0.630)	(0.420)	(0.260)	(0.879)	(0.044*)	(0.467)	(0.319)	(0.432)
$SPRD$	0.143	-0.024	0.189	0.290	-0.385	-0.03	-0.001	0.013	0.004	-0.031
	(0.955)	(0.956)	(0.698)	(0.803)	(0.545)	(0.907)	(0.969)	(0.652)	(0.877)	(0.209)
$EMPL$	-89.819	125.535	-40.638	-119.069	-82.473	1.697	8.666	-0.092	0.430	-0.203
	(0.848)	(0.063)	(0.589)	(0.645)	(0.539)	(0.686)	(0.019*)	(0.980)	(0.941)	(0.970)
$RECES$						0.030	0.001	0.006	0.030	0.077
						(0.079)	(0.938)	(0.791)	(0.144)	(0.001**)
$CRECRU$	-3.510	-0.806	-0.966	-2.081	0.154					
	(0.188)	(0.012*)	(0.023*)	(0.138)	(0.853)					
$constant$	106.831	19.745	18.189	41.331	27.472	4.097	2.825	3.097	2.812	2.359
	(0.000***)	(0.000***)	(0.023*)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)
Sample Size	154	154	154	154	154	154	154	154	154	154
$adjR^2$	0.787	0.617	0.769	0.753	0.789	0.774	0.606	0.741	0.719	0.786
relative weight of $MICS$	7.03%	6.24%	1.66%	7.63%	14.51%					
relative weight of $log(MICS)$						7.59%	6.78%	1.47%	8.56%	15.21%

This table shows the coefficients and p-values of multivariate regressions in Equation (6). Columns 2 to 5 report results from a regression model same as that in Table 7, except for replacing $RECES$ with the newly defined credit crunch indicator ($CRECRU$). Columns 6 to 10 record results from a regression model that takes logged national and regional HBA measures as the dependent variable, as well as logged $MICS$ and logged PMI as replacements for the original series. p-values are based on Newey-West adjusted standard errors, and lead to no qualitatively different results in hypothesis tests from those based on non-adjusted p-values. *, ** and *** represent significance at 5%, 1% and 0.1% levels, respectively.

5.4.4 Bootstrapped p-values

Up to this point, all regression-coefficient tests have been based upon Newey-West Heteroskedasticity and Autocorrelation-Consistent (HAC) standard errors.[17] Nevertheless, the literature has long argued that such HAC corrections hold asymptotically, and therefore may be of questionable validity with finite samples, and especially with relatively small sample sizes. For instance, Goetzmann and Jorion (1993) show that the adoption of bootstrap and HAC corrections lead to opposite inferences regarding whether factors such as dividend yields can be used to forecast future stock prices. And as an anonymous reviewer has pointed out, although the sample size in the present study is technically acceptable, it can be treated as 'small'. Responding to these considerations, in this section we check the robustness of these Newey-West based inferences with bootstrapped p-values.

The bootstrap procedure is implemented as follows:

(a.) For each estimation of Equation (6), coefficients, residuals, and test statistics are stored.

(b.) Residuals are drawn with replacement to generate a bootstrapped artificial residual series.[18]

(c.) For each targeted regressor ($Factor^i$), a pseudo series of the dependent variable is generated under the null hypothesis that a zero coefficient is associated with $Factor^i$. Estimated coefficients from step (a) are used for all other regressors, including the constant, in generating the pseudo dependent variable.

(d.) The pseudo dependent variable is regressed on all regressors. Coefficients and the t-statistic of $Factor^i$ are recorded.

(e.) Steps noted above are repeated 9999 times. The bootstrap sample size is chosen so that $\alpha(1 + B)$ becomes an integer, making the simulation closer to be exact, where α is the significance level and B is the bootstrap sample size (MacKinnon (2006)).

[17] Following Newey and West's original recommendation, the bandwidth parameter is set to $4(T/100)^{2/9}$.

[18] In order to capture the time-series feature of business formation within a contiguous period, e.g. 12 months, we also conduct this step using a blocked-bootstrap approach, where residuals are drawn in blocks of length 12. The last draw is truncated to fit the required sample size. This parallel approach leads to no differences in our statistical-inference results.

(f.) The empirical sampling distribution of the t-statistic under the null hypothesis that $Factor^i$ shows no predictive power in business formation is then obtained by pooling together the 9999 t-statistic values from step (e).

(g.) Reject the null hypothesis at the 5% level if its test statistic t^i from step (a) falls outside the 95% quantile of the empirical sampling distribution.

Results are shown in Tables 12 and 13.

Some small differences are evident between the bootstrapped p-values in Table 12 and the Newey-West p-values reported in Table 5, but none that impinge upon the four principal findings discussed in Section 5.2. With bootstrapped p-values, PMI remains a significant predictor of *BA* at a national level as well as in the Northeast and the Midwest. But whereas with Newey-West p-values PMI is not a significant predictor of *BA* in the South and West regions, with bootstrapped p-values it becomes a significant predictor of *BA* in these regions. With Newey-West p-values, the $RECES$ variable is significant nationally as well as in each individual region. With bootstrapped p-values, the $RECES$ variable is also significant nationally, but only in three out of four individual regions. Finally, whereas the regression constant is non-significant for national-level *BA* under Newey-West p-values, it becomes significant with bootstrapped p-values. Most importantly, the sentiment variable $MICS$ is significant at the 0.1% level nationally and for each sub-national region under both Newey-West and bootstrapped p-values.

Bootstrapped p-values in the decomposition by payroll propensity (Table 13) confirm the robustness of the Newey-West based findings in Table 6 and which are discussed in Section 5.3. The only difference is that according to bootstrapped p-values, the significance of $RECES$ in predicting national-level *LBA* is driven entirely by *LBA* in the West region. All of the key distinctions between *LBA* and *HBA* survive the standard errors robustness check.

Table 12: Coefficients of multivariate regressions and bootstrapped p-values: BA.

$Factor^i$	$BusinessFormation$ measures				
	BA national	BA Northeast	BA Midwest	BA South	BA West
$MICS$	22.717	2.632	2.478	11.901	5.621
	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)
$INFL$	-11.849	-133.536	24.158	-37.789	-173.960
	(0.984)	(0.066)	(0.763)	(0.878)	(0.148)
$T30R$	-1604.729	-61.596	10.791	-1463.090	-48.439
	(0.417)	(0.814)	(0.971)	(0.161)	(0.917)
$CONS$	-239.843	13.443	-47.277	-132.112	-57.211
	(0.669)	(0.869)	(0.582)	(0.621)	(0.655)
$PROD$	325.596	53.152	56.422	152.881	101.775
	(0.197)	(0.166)	(0.141)	(0.207)	(0.076)
PMI	2.191	0.301	0.466	0.931	0.431
	(0.010*)	(0.011*)	(0.000***)	(0.038*)	(0.029*)
$SPR3$	5351.224	640.710	1014.622	2138.016	1044.658
	(0.313)	(0.395)	(0.202)	(0.399)	(0.384)
$SPR10$	80.774	-6.314	8.828	38.279	-5.650
	(0.336)	(0.612)	(0.485)	(0.367)	(0.765)
$SPRD$	709.548	119.651	257.126	281.123	80.729
	(0.454)	(0.359)	(0.064)	(0.553)	(0.701)
$EMPL$	-487.110	152.839	-210.155	-251.410	-112.917
	(0.823)	(0.598)	(0.500)	(0.809)	(0.810)
$RECES$	28.884	3.213	3.921	13.315	7.990
	(0.030*)	(0.065)	(0.041*)	(0.047*)	(0.010**)
$constant$	102.787	21.965	12.291	42.080	29.930
	(0.046*)	(0.001**)	(0.090)	(0.105)	(0.010**)
Sample Size	154	154	154	154	154
$adjR^2$	0.578	0.527	0.425	0.612	0.618
relative weight of $MICS$	62.4%	54.1%	50.8%	64.7%	64.0%

This table shows the coefficients and p-values of multivariate regressions in Equation (6), with national and regional BA measures as the dependent variable. P-values are calculated from bootstrapped empirical distributions of the t-statistic under the null hypothesis of insignificance. *, ** and *** represent significance at 5%, 1% and 0.1% levels, respectively.

Table 13: Coefficients of multivariate regressions and bootstrapped p-values: *LBA* and *HBA*.

	$BusinessFormation$ measures									
$Factor^i$	*LBA* national	*LBA* Northeast	*LBA* Midwest	*LBA* South	*LBA* West	*HBA* national	*HBA* Northeast	*HBA* Midwest	*HBA* South	*HBA* West
$MICS$	20.751	2.506	2.709	10.916	4.621	1.892	0.128	0.237	0.965	1.054
	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.001**)	(0.167)	(0.020*)	(0.0.001**)	(0.000***)
$INFL$	-566.567	-82.877	-79.304	-296.796	-21.584	-12.463	-48.061	40.508	86.574	-36.952
	(0.225)	(0.189)	(0.298)	(0.199)	(0.835)	(0.923)	(0.065)	(0.090)	(0.182)	(0.330)
$T30R$	-8413.977	-930.461	-1229.345	-4403.655	-1855.400	6863.378	867.847	1245.675	2956.309	1787.686
	(0.000***)	(0.001**)	(0.001***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)
$CONS$	-147.875	2.405	-38.800	-74.286	-41.285	-85.386	6.663	-8.979	-54.008	-39.412
	(0.748)	(0.972)	(0.602)	(0.755)	(0.690)	(0.602)	(0.836)	(0.766)	(0.493)	(0.382)
$PROD$	286.671	38.427	48.821	126.836	54.970	117.818	14.121	16.678	46.110	36.707
	(0.197)	(0.200)	(0.166)	(0.234)	(0.240)	(0.103)	(0.313)	(0.210)	(0.198)	(0.070)
PMI	2.120	0.323	0.458	0.881	0.485	-0.054	-0.217	-0.006	0.018	-0.037
	(0.012*)	(0.004**)	(0.001**)	(0.037*)	(0.009**)	(0.743)	(0.449)	(0.832)	(0.825)	(0.401)
$SPR3$	5595.971	744.677	1020.886	2563.669	1501.242	-1171.130	-105.219	-116.520	-582.145	-365.977
	(0.214)	(0.234)	(0.160)	(0.248)	(0.127)	(0.431)	(0.716)	(0.649)	(0.405)	(0.387)
$SPR10$	-3.982	-0.856	-4.019	4.475	11.818	-1.505	-5.928	2.708	12.357	-4.209
	(0.957)	(0.934)	(0.743)	(0.904)	(0.466)	(0.949)	(0.208)	(0.545)	(0.305)	(0.533)
$SPRD$	802.699	121.680	247.886	284.545	133.591	-16.703	-2.263	18.369	18.782	-60.761
	(0.375)	(0.310)	(0.087)	(0.534)	(0.489)	(0.938)	(0.954)	(0.614)	(0.859)	(0.307)
$EMPL$	-568.869	-13.877	-202.897	-240.428	-150.001	217.763	162.524	7.302	25.623	13.106
	(0.777)	(0.959)	(0.537)	(0.813)	(0.727)	(0.619)	(0.044*)	(0.930)	(0.909)	(0.906)
$RECES$	25.891	3.249	3.825	12.380	6.487	2.026	-0.066	-0.021	0.607	1.632
	(0.043*)	(0.057)	(0.063)	(0.057)	(0.019*)	(0.383)	(0.873)	(0.962)	(0.618)	(0.012*)
$constant$	6.559	2.949	-4.215	4.603	1.856	102.978	19.011	17.293	39.215	27.152
	(0.895)	(0.646)	(0.583)	(0.846)	(0.859)	(0.000***)	(0.000***)	(0.000***)	(0.000***)	(0.000***)
$adjR^2$	0.645	0.611	0.576	0.665	0.649	0.784	0.606	0.758	0.745	0.797
relative weight of $MICS$	48.0%	45.7%	37.5%	50.6%	48.2%	7.1%	6.4%	1.3%	7.9%	14.5%

This table shows the coefficients and p-values of multivariate regressions in Equation (6), with national and regional *HBA* and *LBA* measures as the dependent variable. P-values are calculated from bootstrapped empirical distributions of the t-statistic under the null hypothesis of insignificance. *, ** and *** represent significance at 5%, 1% and 0.1% levels, respectively.

6 Conclusion

This study brings both new data and new conceptual apparatus to bear upon entrepreneurial business formation. The US Census Bureau's weekly business formation statistics, which we aggregate up to the monthly frequency, can be a rich resource for future entrepreneurship research. In this study we exploit the overlap between entrepreneurship and finance to justify the use of behavioral finance concepts such as fundamentals-focused information processing as well as its converse, which is susceptible to classifying noise as signal and vice versa. In this second category, widespread mood or affect – i.e. sentiment – can influence decision making.

Our results show that broad business formation is jointly determined by economic fundamentals and consumer sentiment. Sentiment, proxied by the Michigan Index of Consumer Sentiment, predicts month-ahead business formation positively and significantly. Thus, sentiment operates as an opportunity-pull factor explaining future business formation. Although two fundamentals variables (the composite purchasing managers' index and the 1-month real US Treasury bill return) show some predictive power in particular regions, one further variable stands out nationally and across all regions: the recession indicator. This variable operates as a necessity-push factor explaining future business formation.

The Census Bureau data allows business formation to be partitioned into high-payroll-propensity and low-payroll-propensity subsets. Separate analysis of these subsets reveals that the aggregate-level results conceal two very different response patterns. High-propensity business formation is mainly driven by fundamentals, with sentiment accounting for a small proportion of explained variation. Low-propensity business formation, on the other hand, is jointly driven by consumer sentiment and fundamentals, with sentiment accounting for close to half of explained variation. Moreover, the short-term interest rate variable (1-month real US Treasury bill return) predicts month-ahead low-propensity business formation *negatively* and month-ahead high-propensity business formation *positively*. The results for high-propensity entrepreneurs are consistent with fundamentals-oriented information processing responding primarily to pull motives. In contrast, the results for low-propensity entrepreneurs are consistent with not solely fundamentals-oriented – i.e.

also noise- and sentiment oriented – information processing responding to both push and pull motives.

The results indicate not only that some entrepreneurs are responding to external-trigger factors while others are responding to internal-trigger factors, but that a subset of the entrepreneur population appears to be responding simultaneously to both internal- and external-trigger factors. An individual's mood or affect is a transient internal-trigger factor, but when individuals' mood or affect is correlated across the economy, it also satisfies the criteria for being an external-trigger factor. Thus sentiment is neither exclusively an internal-trigger factor nor exclusively an external-trigger factor. The question of whether the *effective* trigger is internal, external, or both, is left for future work to enrich the conception of sentiment employed in both behavioral finance and entrepreneurship.

References

Amit, R., and E. Muller. 1995. "'Push' and 'pull' entrepreneurship." *Journal of Small Business Entrepreneurship* 12(4): 64–80.

Ardagna, S., and A. Lusardi, 2010. "Explaining international differences in entrepreneurship: The role of individual characteristics and regulatory constraints." In *International Differences in Entrepreneurship*, edited by J. Lerner and A. Schoar, 17–62. Chicago: NBER and Chicago University Press.

Baron, R.A. 2000. "Psychological perspectives on entrepreneurship: Cognitive and social factors in entrepreneurs' success." *Current Directions in Psychological Science* 9(1): 15–18.

Baker, M., and J. Wurgler. 2007. "Investor sentiment in the stock market." *Journal of Economic Perspectives* 21(2): 129–151.

Bayard, K., E. Dinlersoz, T. Dunne, H. Haltiwanger, J. Miranda, and J. Stevens. 2018. "Early-stage business formation: An analysis of applications for employer identification numbers." Finance and Economics Discussion Series 2018-015. Washington, DC: Board of Governors of the Federal Reserve System.

Bergmann, H., S. Mueller, T. Schrettle. 2014. "The use of Global Entrepreneurship Monitor data in academic research: A critical inventory and future potentials." *International Journal of Entrepreneurial Venturing* 6(3): 242–276.

Brown, G.W., and M.T. Cliff. 2005. "Investor sentiment and asset valuation." *Journal of Business* 78, 405-440.

Cleveland, R.B., W.S. Cleveland, J.E. McRae, and I. Terpenning. 1990. "STL: A seasonal-trend decomposition procedure based on loess." *Journal of Official Statistics* 6(1): 3–33.

Decker, R., Haltiwanger, J., Jarmin, R. and J. Miranda. 2014. "The role of entrepreneurship in US job creation and economic dynamism." *Journal Economic Perspectives* 28(3): 3–24.

Evans, D.S. and J. Jovanovic. 1989. “An estimated model of entrepreneurial choice under liquidity constraints.” *Journal of Political Economy* 97(4): 808–827.

Fairlie, R.W., and F.M. Fossen. 2019. “Defining opportunity versus necessity entrepreneurship: Two components of business creation.” NBER Working Paper 26377.

Frese, M., and M.M. Gielnik. 2014. “The psychology of entrepreneurship.” *Annual Review of Organizational Psychology and Organizational Behavior* 1: 413–438.

Gilad, B., and P. Levine. 1986. “A behavioral model of entrepreneurial activity.” *Journal of Small Business Management* 24(4): 45–53.

Goetzmann, W.N., and P. Jorion, 1993. “Testing the Predictive Power of DividendYields.“ *The Journal of Finance* 48: 663–679.

Gough, J.W. 1969. *The Rise of the Entrepreneur.* New York: Schocken Books.

Hurst, E., and B.W. Pugsley. 2011. “What do small businesses do?” *Brookings Papers on Economic Activity* 2: 73–118.

Jarmin, R.S., and J. Miranda. 2002. “The longitudinal business database.” Working Paper Number CES-02-17, Center for Economic Studies, U.S. Census Bureau, Washington, DC. https://www.census.gov/library/working-papers/2002/adrm/ces-wp-02-17.html

Johnson, J.W. 2000. “A heuristic method for estimating the relative weight of predictor variables in multiple regression” *Multivariate Behavioral Research* 35: 1–19.

Klapper, L., R. Amit, M.F. Guillén, and J.M. Quesada. 2010. “Entrepreneurship and Firm Formation Across Countries.” In *International Differences in Entrepreneurship*, edited by J. Lerner, and A. Schoar, 129–158. Chicago: NBER and University of Chicago Press.

Lemmon, M., Portniaguina, E., 2006. “Consumer confidence and asset prices: some empirical evidence.” *Review of Financial Studies* 19(4): 1499–1529.

Licht, A.N., and J.I. Siegel. 2006. “The social dimensions of entrepreneurship.” In *Oxford Handbook of Entrepreneurship*, edited by M Casson, B. Yeung, A. Basu, and N. Wadeson, 511–539. Oxford: Oxford University Press.

Lucas, R.E. Jr., 1978. “Asset prices in an exchange economy” *Econometrica* 46(6): 1429–1445.

MacKinnon, J.G., 2006. ”Bootstrap Methods in Econometrics”. *Economic Record* 82, S2–S18.

Qiu, L., Welch, I., 2006. “Investor sentiment measures.” *NBER Working Paper* 10794.

Salmerón, R. C. García, and J. García. 2018. “Variance inflation factor and condition number in multiple linear regression.” *Journal of Statistical Computation and Simulation* 88(12): 2365–2384.

Schoar, A. 2010. “The divide between subsistence and transformational entrepreneurship.” In *NBER Innovation Policy and the Economy, Volume 10*, edited by J. Lerner and S. Stern: 57–81. Chicago: NBER and Chicago University Press.

Shiller, R.J., 1999. “Measuring bubble expectations and investor confidence.” *NBER Working Paper* 7008.

Schumpeter, J.A. 1911. *Theorie Der Wirtschaftlichen Entwicklung.* Leipzig: Duncker & Humblot.

Sternberg, R., and S. Wennekers. 2005. “Determinants and effects of new business creation using Global Entrepreneurship Monitor data.” *Small Business Economics, Special Issue* 24(3): 193–203.

Tonidandel, S., and J.M. LeBreton. 2011. “Relative Importance Analysis: A Useful Supplement to Regression Analysis” *Journal of Business and Psychology* 26(1): 1–9.

U.S. Census Bureau. 2020. “Business formation statistics, second quarter 2020.” Release Number CB20-106, U.S. Census Bureau, Washington, DC. https://www.census.gov/econ/bfs/index.html .